

Thu Oct 14 07:57:23 1999

us-09-075-375-1.rge

Clust

Page 1

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 1999, 02:48:22 (Search time 2132.21 Seconds
2954.771 Million cell updates/sec)

US-09-075-375-1

Title: 1991
Perfect score: 1
Sequence: 1 tcttggtcgaatttacc.....aattttgctaaaaaaa 1981

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database:

GenBank: 1: gb_ba1: 2: gb_ba2: 3: gb_om: 4: gb_ov: 5: gb_pat: 6: gb_ph: 7: gb_p11: 8: gb_p12: 9: gb_p13: 10: gb_p14: 11: gb_p15: 12: gb_p16: 13: gb_p17: 14: gb_p18: 15: gb_p19: 16: gb_p20: 17: gb_p21: 18: gb_p22: 19: gb_p23: 20: gb_p24: 21: gb_p25: 22: gb_p26: 23: gb_p27: 24: gb_p28: 25: gb_p29: 26: gb_p30: 27: gb_p31: 28: gb_p32: 29: gb_p33: 30: gb_p34: 31: gb_p35: 32: gb_p36: 33: gb_p37: 34: gb_p38: 35: gb_p39: 36: gb_p40: 37: gb_p41: 38: gb_p42: 39: gb_p43: 40: gb_p44: 41: gb_p45: 42: gb_p46:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1759	88.8	1760	8	LSU31462	U31462 Lactuca sat

2	643.8	32.5	1589	8	NTU34817
3	611.8	30.9	1555	8	ATU44133
4	406.4	20.5	13840	8	AC003981
5	404.8	20.4	2884	8	ATU39452
6	143.2	7.7	89493	8	AC006932
7	78	3.5	7218	5	166494
8	52.6	2.3	108409	17	AF083424
9	51.8	2.4	118595	9	HSJ288L1
10	51.4	2.4	153053	34	HSJ288L1
11	50.4	2.5	11228	10	HSJ288L1
12	49.2	2.5	13095	10	HSJ288L1
13	49.2	2.5	194410	35	AC005140
14	49.2	2.5	219200	35	AC006219
15	49.2	2.5	143769	42	AC004914
16	49	2.5	100000	9	AP000067
17	48.6	2.4	11223	12	RNU07886
18	48.2	2.4	113843	11	AC005033
19	47.8	2.4	130508	17	AF005370
20	47.8	2.4	130508	34	AC004965
21	47.8	2.4	130508	35	AC006329
22	47.8	2.4	207984	35	AC006544
23	47.8	2.4	234523	42	AC005065
24	47.6	2.4	90832	11	AC007461
25	47.6	2.4	180385	42	AC007461
26	47.2	2.4	13674	36	U00691
27	47.2	2.4	13729	36	U00796
28	46.8	2.4	194985	12	AC002406
29	46.8	2.4	194985	12	MMHC438N12
30	46.6	2.4	4122	37	DDU50042
31	46.6	2.3	161475	9	HS164C20
32	46.4	2.3	3200	9	HUMGAL
33	46.4	2.3	39229	11	AC004755
34	46.4	2.3	34450	11	HUAC003049
35	46.4	2.3	4238	12	RNCNG41
36	46.4	2.3	142228	34	AC004123
37	46.2	2.3	6372	35	DDICNPA
38	46	2.3	202233	11	PFMAL364
39	45.8	2.3	65691	36	PFSC03005
40	45.8	2.3	5600	36	PFSC03005
41	45.6	2.3	119857	34	AC006782
42	45.6	2.3	144514	34	HS1018D12
43	45.6	2.3	194410	35	AC005140
44	45.6	2.3	5197	36	PFSC03053
45	45.4	2.3	394	11	HSFRAH1

ALIGNMENTS

RESULT	1	PLN	23-JUL-1996
LSU31462	1760 bp	MRNA	
LOCUS	Lactuca sativa. Violaxanthin de-epoxidase (VDEL) precursor, MRNA, complete c.s.		
DEFINITION	U31462		
ACCESSION	U31462		
VERSION	U31462.1	GI:1438874	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

1 (bases) to 1760
Bugos, R.C. and Yamamoto, H.Y.
Molecular cloning of violaxanthin de-epoxidase from romaine lettuce and expression in *Escherichia coli* (13), 6320-6325 (1996)
Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6320-6325 (1996)
2 (bases) to 1760
Bugos, R.C.
Direct submission
Submitted (12-JUL-1995) Plant Molecular Physiology, University of

FEATURES Hawaii, 3190 Maile Way, Honolulu, HI 96822, USA
 Source Location/Qualifiers
 1.1760
 /organism="Lactuca sativa"
 /strain="Romaine"
 /db_xref="taxon:4236"

transit_peptide 235..609
 /gene="VDE1"

CDS 235..1656
 /gene="VDE1"

/codon_start=1

/product="violaxanthin de-epoxidase precursor"

/protein_id="AAC49373.1"

/db_xref="PIR:G1438875"

/db_xref="GI:1438875"

/translation="MALSLHTFLVLCKEALNLVARSPCNRRHSGPPNITIMKIR
 MNQVNSFRFTSYKTSFSDSHCKDKSCISIDTSPEEIOREIDRKRGNTLLEKQ
 TNRPIQLAIVLCFTVIVPRVADALFKACILKECTELAKIPIANPCANVACIO
 FNKDFSGKWIITSGNLTEDAFDQDLHEFHNENKCVKPSVDGEPPVDRNAVON
 QTEVODPDLGALYNHNEFLHYODWLTLSOLENKKDDYIEVYGRNDAMDYGG
 SVIYRSPILPEIIPNLOKAKSVGRFNNFTIDNSGPEPLVERLEKTEBEK
 LLIKAEVLEEEVEKEKVEKVDTEMTLFLRLESGFKEHOOEENFVRELSEKEELN
 ELQMEATEVEKEKFGKALPLRKLK"

gene 235..1656
 /gene="VDE1"

mat_peptide 610..1653
 /gene="VDE1"

/product="violaxanthin de-epoxidase"

BASE COUNT 557 a 319 c 387 g 497 t

ORIGIN

Query Match 88.8% Score 1759; DB 8; Length 1760;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgggttcgaattttaccaccacaagtttgcctaccataattgggataaggctc 60
 DB 1 ttgggttcgaattttaccaccacaagtttgcctaccataattgggataaggctc 60
 QY 61 aattccctgtacaattttccattctctccctccgcaacatatatactacgc 120
 DB 61 aattccctgtacaattttccattctctccctccgcaacatatatactacgc 120
 QY 121 cacttcgaagcctacaattgtttgaaaaaagcagaattttcaagaagcgaagataa 180
 DB 121 cacttcgaagcctacaattgtttgaaaaaagcagaattttcaagaagcgaagataa 180
 QY 181 taagcttcaagtaactccgactcgcagtggtgctttgaaagcacaacacgtgcctatgct 240
 DB 181 taagcttcaagtaactccgactcgcagtggtgctttgaaagcacaacacgtgcctatgct 240
 QY 241 ctcttccttcaactgtatctctctcgaagaagagccctcaattatatactgcaagatca 300
 DB 241 ctcttccttcaactgtatctctctcgaagaagagccctcaattatatactgcaagatca 300
 QY 301 ccatgttaatgaaggcttcacagagtggaacacccctcccaacaataatcatgataa 360
 DB 301 ccatgttaatgaaggcttcacagagtggaacacccctcccaacaataatcatgataa 360
 QY 361 attcaatcaacaatgatatttaattcttccggtgttttaccattataagaagaat 420
 DB 361 attcaatcaacaatgatatttaattcttccggtgttttaccattataagaagaat 420
 QY 421 tcttctcagaatcgaacatcgaagataatcgaatataatgaatcgaatcgaatcgaat 480
 DB 421 tcttctcagaatcgaacatcgaagataatcgaatataatgaatcgaatcgaatcgaat 480
 QY 481 tttaaggaataacagaatttgatctcaaaaggcgcagacttgatcttgaaagcaaa 540
 DB 481 tttaaggaataacagaatttgatctcaaaaggcgcagacttgatcttgaaagcaaa 540

QY 541 tggagacaattcattacataattgctatcgatctgattgtttcacaattgttatgtccaca 600
 DB 541 tggagacaattcattacataattgctatcgatctgattgtttcacaattgttatgtccaca 600
 QY 601 gtgtatgcgttgatgctctcttaaaacttgctgttttaccccaagaagaatcgaatgag 660
 DB 601 gtgtatgcgttgatgctctcttaaaacttgctgttttaccccaagaagaatcgaatgag 660
 QY 661 ctgtcaaaatgatatgaacccaacatctgtgagcgaacagcttcggtcccaagacttc 720
 DB 661 ctgtcaaaatgatatgaacccaacatctgtgagcgaacagcttcggtcccaagacttc 720
 QY 721 aacaaatcgtccctcgaagagccgaatgtcagaataaagtgtgactgttgcgaagaacgt 780
 DB 721 aacaaatcgtccctcgaagagccgaatgtcagaataaagtgtgactgttgcgaagaacgt 780
 QY 781 gtgttgagcccaatcgaacaggtgtggtgttccggaagaatgtgtgccccggaatcg 840
 DB 781 gtgttgagcccaatcgaacaggtgtggtgttccggaagaatgtgtgccccggaatcg 840
 QY 841 gattgagtgatattcccggttccggaatcgtaatgagagtggttcaaatattcaatgaa 900
 DB 841 gattgagtgatattcccggttccggaatcgtaatgagagtggttcaaatattcaatgaa 900
 QY 901 gacttaagtggagagtggtatataaagaatgttcaatcctacatttgatgacattgaa 960
 DB 901 gacttaagtggagagtggtatataaagaatgttcaatcctacatttgatgacattgaa 960
 QY 961 tgcacaattcagagttcattatataatgaagaatgataaactgtttggaacttaagaatg 1020
 DB 961 tgcacaattcagagttcattatataatgaagaatgataaactgtttggaacttaagaatg 1020
 QY 1021 ataaaaacttgagagtggtgttcttaccatgactcgtgtgcaaacattgttcaaat 1080
 DB 1021 ataaaaacttgagagtggtgttcttaccatgactcgtgtgcaaacattgttcaaat 1080
 QY 1081 ccaagcttctcctgagagcattatataatcagagagtgatcttcttaccacaaatgag 1140
 DB 1081 ccaagcttctcctgagagcattatataatcagagagtgatcttcttaccacaaatgag 1140
 QY 1141 tggatcatattatcttcccaatcgaagaacaaacccgagatataatctgatactac 1200
 DB 1141 tggatcatattatcttcccaatcgaagaacaaacccgagatataatctgatactac 1200
 QY 1201 cgaagctcgaacacgaatgagagtgatgagtggtgtgctgagatcctaaacccgagagc 1260
 DB 1201 cgaagctcgaacacgaatgagagtgatgagtggtgtgctgagatcctaaacccgagagc 1260
 QY 1261 aacactcccggaatcgatcattcccaacacgaagaacaaacccgagagc 1320
 DB 1261 aacactcccggaatcgatcattcccaacacgaagaacaaacccgagagc 1320
 QY 1321 tttaacaattatataacacgaacagagatgtgtgagcctgagcctcatgtgtggaag 1380
 DB 1321 tttaacaattatataacacgaacagagatgtgtgagcctgagcctcatgtgtggaag 1380
 QY 1381 ctgtggaagaacagcgaagagcgaagaatgtgtgataaagaagctgttaagaatgaa 1440
 DB 1381 ctgtggaagaacagcgaagagcgaagaatgtgtgataaagaagctgttaagaatgaa 1440
 QY 1441 gaaagagtgtaaaagagtggtggaagagtgtaagaatactgagatcttcaagag 1500
 DB 1441 gaaagagtgtaaaagagtggtggaagagtgtaagaatactgagatcttcaagag 1500
 QY 1501 ttgctttaaagggttttaaggagttgcaacaagaatgtaagaagatgtgtgaggaatgag 1560
 DB 1501 ttgctttaaagggttttaaggagttgcaacaagaatgtaagaagatgtgtgaggaatgag 1560
 QY 1561 aaagaagaagaagaatcttgatgaacttcaaaatggaagagcagactggaatgaaagctt 1620
 DB 1561 aaagaagaagaagaatcttgatgaacttcaaaatggaagagcagactggaatgaaagctt 1620
 QY 1621 ttgggagcgcgttaccgattaggaacttagataaatttcgatgtatgtattcagaacaat 1680

	Best Local Similarity	71.8%	Pred. No. 1,9e-123:		Matches	862:	Conservative	0:	Mismatches	327:	Indels	12:	Gaps	1:
QY	483	tgaagaatacacaagaattgcatctcaaaaggagccatgacttgcttcttgaagaacatg	542											
Db	317	TAAAGCAAAATTCAGATGCAAAATTCCAAAAGATGTACTGACTTGAGTTGGAGAAATG	376											
OY	543	gagacaattc*taaatgtctatgcttatgtttgtgcacattgttactcgttccagagt	602											
Db	377	GGGCCAATTTCGCAAAACACCAATTTGACTCATATTCATTGTGTAGTTGGCTTCAAAAGC	436											
OY	603	tgaagccgttatgtctcttaaaacttgctgctgtttaccctcaagaatgcagatgaaact	662											
Db	437	TCATCGGTTGATGCTCTCAAGACTGTGTACTTCTTCTTACGAAAGAGTGCAGTTAGACT	496											
OY	663	tgcacaatgtatagcacaaccatctgttggcgcaaacgcttgctctgtctacagacttgc	722											
Db	497	TCCGAAGTCAATTCGAACCTTGATGTGACGTAAATGTGGCTGTCTCCACACTTGC	556											
OY	723	caatgctctgcagagaccgaatgtcagataaaatgtgtgtgacttgcttgcgaacaagt	782											
Db	557	CATTAAGCTGACGAACGAGATGTCAAGTAAATGTGGTGGATTTGTTTGAACAGTGT	616											
OY	783	gattgagccaatccaacgaagtgtgagtttcccgaagaatgtgtgcccgaatccga	842											
Db	617	CGTAGACGATTCAATGAGTGTGCAAGTCTCCGAAAGAAATGTGTACTCTGTAAATCTGA	676											
OY	843	tgtgtgtgaattcccggttccggatcgtaatgcaagtgtgtcaaatcttcaactgaaga	902											
Db	677	TGTTGGTGAATTCCTGTACTGATCCAGTGTCTTCTTCAGAAAGTTTGACATGAAGA	736											
OY	903	ctttaagtgggaagtgtgtatatacaaatgtgtttaaactccatcttgatgacttgat	962											
Db	737	TTTATGGGGAAATGTTGTCATTACTGCGCGTTGAAATCCCACTTTTGATGCTTTTGATG	796											
OY	963	tcaacttcattgatttcataatggaataatgaatcttcttggaaacttaaatgtgcgat	1022											
Db	797	CCAAATGATAGTTCCTACACGAAGAAACAAACTTTGGGGATTTATCTTGAGAAAT	856											
OY	1023	aaaaacttgaatgtgtgttcttcttactcgaatcgtctgtgcgaacattgttcaagatcc	1082											
Db	857	ACGTACACCTATGAGAGATTTTACTGCATAGCGGTGCAGAAATTCGTGCAGATCC	916											
OY	1083	agatcttcctgtgagcaactttatctatctgacaatgtgttcttcaactacaagatgact	1142											
Db	917	AAAGTATCCGGGATACCTTCATTCATGTGATGATGATCTCTCTACCAAGATGACAG	976											
OY	1143	gtaactatata*cttccaactcgaaacaaaccgatgattacatatctgttactaacg	1202											
Db	977	GTATATTTTGT*ATCCAAAGTGAAGAAATAGCCAGAGATTTACATATTTGTGTACTATA	1036											
OY	1203	aggttcgaacgaacgaatggatggatagatgggtgggtcgtgtatctaacaccggaagcccgac	1262											
Db	1037	GGGCAGAAATGATGCAATGGGATGGATATGGTGTCTGTACTTACACAGAAAGTGACAT	1096											
OY	1263	acgtcccggaatgatcatcccaaacactcaaaaagcagcccaatccgttggctcgagact	1322											
Db	1097	TTTTCCTGAAAT*ATTATACCGGAGTGTGCAAACCCACACTCAAAAAGTGGGCTGATTT	1156											
OY	1323	taacaattcttaacaaccgcgaataatgtgtgtggctctagctccatgtgtgtga												

Db 1325 GTTGAAGGTTTAAAGAGCTCCACAGATGAGAGAACTTTAAAGAGCTGACCA 1384
 QY 1553 agaagaagaagaatttcgaataactcaaatggaagcgaactgaattgaagcttt 1622
 Db 1385 AGAAGAAATGATGTTTGGATGACTTAAATGAGCAACTGAGTAGAATAACTTTT 1444
 QY 1623 tggcgcgcgttcacgattgagaactagataaattcgaattgattgaagaactat 1682
 Db 1445 TGGGCGTGCTTTACCAATTAAGAAATTAAGTAAGTATTTTAAACTATCAACATAT 1504
 QY 1683 a 1683
 Db 1505 A 1505
 RESULT 3
 ATU4133
 LOCUS 1555 bp mRNA PLN 15-JUN-1998
 DEFINITION Arabidopsis thaliana violaxanthin de-epoxidase precursor (AVDE1)
 ACCESSION U44133
 NID g1465734
 VERSION U44133.1 GI:1465734
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis
 1 (bases 1 to 1555)
 Bugos, R.C.; Hieber, A.D. and Yamamoto, H.Y.
 Xanthophyll cycle enzymes are members of the lipocalin family, the
 first identified from plants
 J. Biol. Chem. 273 (25), 15321-15324 (1998)
 MEDLINE 98288256
 2 (bases 1 to 1555)
 Bugos, R.C. and Yamamoto, H.Y.
 Direct Submission
 Submitted (02-JAN-1996) Robert C. Bugos, Plant Molecular
 Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI
 96822, USA
 FEATURES
 source
 location/Qualifiers
 1..1555
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 1..1555
 /gene="AVDE1"
 45..1433
 /gene="AVDE1"
 /codon_start=1
 /product="violaxanthin de-epoxidase precursor"
 /protein_id="AAC50032.1"
 /db_xref="PIR:G1465735"
 /translation="MAVATHCFTSPCHRIIPFSSDGCIRLITTKRINGLFKLIL
 CAPIIYPSADAVDALKTCACILKGRIFLACINPACANVACLOTNNRPDEECCO
 TSGDLFENSVDDEFNEPCARCKVCPRKSDGEPFAPDPFVAVQWNTSDNGKWI
 VLYNDNEYLIHODMTILSKIKENKPEYIFVYRGNDAMGCGAATVYRRSVLP
 NSTIRDELKAKKSGIRGFSTFIRTDNCGEPALVBERIEKTVGEGRIIVKEVIEE
 EYKEVEKVGKTEMTLFLORLAEFNELEKODEENFVLELSKEKEFLDETKMASEVER
 LERKALPIRKVR"
 45..383
 /gene="AVDE1"
 384..1430
 /gene="AVDE1"
 /product="violaxanthin de-epoxidase"
 BASE COUNT 485 a 285 c 376 g 409 t
 ORIGIN

Query Match 30.9%; Score 611.8; DB 8; Length 1555;
 Best Local Similarity 72.5%; Pred. No. 7.3e-117;
 Matches 806; Conservative 0; Mismatches 302; Indels 3; Gaps 1;
 QY 550 ttctacaattgctatcgtatgttggttgacattgttgcgttcccaagttgagcc 609
 Db 324 TTGCTAAACTCGTGCGTGTATTACCTGCGGCTTCCTATTGTTCCATCTGCAGATGA 383
 QY 610 gttagtcctctaaactgattgctgttctactcaagaatgcagattgagcttgcaaa 669
 Db 384 GTTGATGCACTTAAACTTGTGATGCTTATTGAAGGATGACGATATGACATCGCAAG 443
 QY 670 tgratagcaaacccactcttgctgacgaacgctgctctctacagcttgcacaactgt 729
 Db 444 TGCATTGCCAACCCCTGCTGTGACGCCAATGTCCGCTTCAGACCTGCAATTAACCGT 503
 QY 730 cctgacgagaccgaatgacataaattgtgtgactgttgcgaacagtggtgagac 789
 Db 504 CCAGATGAACCGAGTGCAGATTAAATGCGGATCTGTTGAGACAGCTTGTGTAT 563
 QY 790 caattcaacgagtgctggttcccgaaagaatggtccccggaatcgatgtggt 849
 Db 564 GAGTTCACAGAGTGTCTGTGCGAATAAAGTGTCTTCAAGAAATGATCTCGGA 623
 QY 850 gaattccggttcgagtgtaagtgatgagtgatgagtgatgagtgatgagtgatgagtg 909
 Db 624 GAATTCCTGCCCCACACCTTGTGTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 683
 QY 910 ggggaagtgatatacaaatggtttaaactcattgattgattgattgattgattgattg 969
 Db 684 GGGAGGGGATCAATTAAGATGCTGTAATCAACCTTGTGTTGTTGTTGTTGTTGTTGTT 743
 QY 970 catgagttcataaggaagatgataaactgattgattgattgattgattgattgattgatt 1026
 Db 744 CATGAGTTCACACAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 803
 QY 1027 acttgatgagtgatttcttactgattgattgattgattgattgattgattgattgattg 1086
 Db 804 ACCCTGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 863
 QY 1087 ctctcggagcattataatcatgacgaatgatttcttactcacaagatgattgattgattg 1146
 Db 864 CAACCTGCTGCTCTCTAATCATGACAGAGTGTGATGATGATGATGATGATGATGATG 923
 QY 1147 atattatcttcccaatgcgaacacccagcatttacttacttacttacttacttacttact 1206
 Db 924 ATCTGTATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 983
 QY 1207 cgaacgacgacatgagtgatagtcggtggtcgtgatttactacccggaagccgacatc 1266
 Db 984 CGAAACGATGCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1043
 QY 1267 ccggaatcgatccatcccaactacaaagaagcgaatccggtggtggtggtggtggtggtg 1326
 Db 1044 CCAATATGCAATTTATACCAAGCTGCAAAAAGCAGCAAAAACATATGACAGAGTGTGATG 1103
 QY 1327 aatttcaacaacccaacatagttggtggtcgtggtccatgtgtggaagaggtgag 1386
 Db 1104 ACATTCATTGAAAGGATATACATGATGATGATGATGATGATGATGATGATGATGATG 1163
 QY 1387 aaaaagcggaagagggcggaagattgtgtgtaaagaagctggaagatgaagaagag 1446
 Db 1164 AAGACAGTGAAGAGGATGAAGATATCTTAAGAGTGAAGAGATGAAGAGAG 1223
 QY 1447 gtgtaaaagagtggaagaggttagagatactgagatgattgttcaagattgtcct 1506
 Db 1224 GTAGAGAAAGAGTGAAGAGTGGTATGAGATGAGATGAGATGAGATGAGATGAGATG 1283
 QY 1507 gaaaggttaagaggtgcaacgaatgaaagaaatttggaggaggtggtgaaagaa 1566
 Db 1284 GAAAGATTTAATGAAGTGAAGCAAGACGAGAGATTTGCTGAGAGATTTAAGTAAAGAA 1343

CDS

ORLIRKREKMDVEKGLVTMEDVEMASQAASDSKSPKFSVSKSLKLNKLN
 KGKKKKNOXASGSDCMLE"
 16782..19871
 /note="similar to receptor-like protein kinase (Ipomoea
 nil) (U77888)"
 /codon_start=1
 /evidence=not_experimental
 /product="F22013.7"
 /protein_id="AAC14033.1"
 /db_xref="PID:g3063445"
 /db_xref="GI:3063445"

CDS

/translation="MAIPRLFIFYIGFALFPVSETPFONSEOEIIIAKSPLEP
 SNNLDMKRPENATTFSELVCHMTGVCHDANGYAKLLSNMKNVSDIOSPS
 LGALDSNNAFESSLPKLSNLTSLKVIDVSNSEFTGFPYGLMAGNGLIHVANSIN
 FSGFLPEDGNATLLEVLDFRGYEGESVSPFNKIKTKLIGLSGNNGKVPKYG
 ELSSLETLILGNGFMGEIPEPEKTLTLQYLDIYAVGLTQITSSGOLKQITLYL
 YONRLTGKLPRLGMSLVLFDLSDNOITGETIPEYGEKLNLOLMLNOLGTLIP
 SKIAPLNLEVLQMONSLMGLSPYHJLQKNSLPKWLVDVSSKLSGDIPLSGICSRNLT
 KLIIPNSESQIPEITSCPLVLRVROKNIHIGSIPAGSDLPMLQHELEKNNLT
 GIPDDIALSTSTIDISENHLSSLSFSPNLOTFTASHNNPQKIPNOIDRP
 SLVIDISFNHPSGIEPIRIASFELVSLNSQLQVGEIPKLAGHMLAVTDLSEN
 SLTGPIPADLGAAPLEMLNVSNFKLDGPIESNMLFPAIDPKLVGNNGCGVADPC
 SKSLASAKGRNPRIHYNHAFGEIIGTIVYAMGMPLAGWITITKMDLSTNARE
 IJCKKPREEPWRIYAFORLCTAGDILSHIKESNTIGCATIGYKRAVMKRPILT
 VAVKLMRSPONDIEDHOEDEDDIIRVNLJGGLHRTVILGLYVHNREEM
 MYEYPMNGNLGTALHSGKDEKPIRLPMLSXINAVGVGGLNHLNDCCPIIHRDIK
 SNNILDSNLEARTADSGAKKMMHKREIVSVAGSYGIAEYGTLLIDERSDYS
 LGVLLLELVTKMPIDSEFSDIDVVEIMRRKVRKSELEVIDASIGADCKHVEEM
 JOLN(22485..22946,23045..23140,23223..23315,23426..23616,
 23701..24112,24456..24569,24699..24856,25371..25724,
 26173..26486,26582..26672,26778..26987,27139..27258,
 27398..27564,27925..28056,28325..28474,28559..28618,
 28675..28756,28805..29002,29146..29275,29686..29741,
 29856..30082,30163..30366,30450..30528,30616..30875,
 30973..31354)
 /note="similar to DNA dependent ATPase and helicase
 (U72937)"
 /codon_start=1
 /evidence=not_experimental
 /product="F22013.8"
 /protein_id="AAC14034.1"
 /db_xref="PID:g3063446"
 /db_xref="GI:3063446"

/translation="WEANESLCKIKLEIKLEKVEIKSEDEMIIIEENREAEVEM
 EVKARDGREBNDVLEENNNQGEKDEMDASRSRSPNDDEBOLSLRDEL
 DLEVCLMDISLGGIYVYLIFLPSIIGKPLSEPIDLSLDLAEKRAEOE
 ALEKSLKSEVEEELAOALRGDELDENAYAMMTFKDEMETIDELTEKAEAOE
 OLDGAGILPLKLEYEISOANGCYTPAKMRAWVOTQYKTEVSLANAERFLTH
 RPKRKHGLLLEEGASGLFEKKLADKAKVBSLAPISFLDMSLNTKVFSEKRDSEVFG
 SKOMASVYLASTPHOAMKLEFPGVNEVIFSLPISFLVLMMLKICLHLIFVA
 PKKLANVSPFIDRLILAVRLKRAVEIEIDASLADFLDADNRELATIEEO
 KTNIRISGIRSFETITTSGLSCFCGVGSKSIYMYHILRLITTLALPKMSM
 TLNHOVKEEDDITCDVLOLRKRRKRRSKOYIRCAENMDSDSYLDGNPTSP
 FAKDOVSPETSTOVHSEVNIENGENSENSDKVTPSHITVADKRDSONPANNE
 RCTLCNVAVAHSHPLEVIVCDKRSIEDRVSVDSLEHCHGCHLDIDR
 TCELCNFASCIRNIGIEYMSPAOSGMOCCSPIDRLTLEKAROKDSIELS
 SDSSSDSSNNVDIDADVNTISSEKSKKIRIIDDAELGKIDTKIIEKARO
 ERLSLIOFSARYKTISSMDVYSIPGAELVEGDHSGYIYVVAEIEGEVVRPS
 ISAKLKHOTVGIREFKMENTIOSISRKKGDKLGLILAHMTGKIDTAKIIEKARO
 RCYDILGKTLITLVNPNVNLNRSEFEKEMSEVRLFEMLGIDVYVAVTAVLFTW
 KKGCVFLMGITINRLSLRGYKDLNAAIGICNARLDGDTIYCEBATIKNTADT
 QALKOVCORRIALTGSPLONNMEYVAVDNCVADNGLSSSEFNRSFNGGSPYD
 LPSNKRRIAYFENCGCFVORMDNVVKDLPRTTVEYISVLSPLORLIYOFLEL
 YGFSDRTERMRKKNFAVYVLAQLNHPGIPOLRSDESKNGRGSIVDIPDCSSD

Query Match 20.5% Score 406.4; DB 8; Length 133840;
 Best Local Similarity 66.3%; Pred. No. 1.1e-74;
 Matches 668; Conservative 0; Mismatches 246; Indels 94; Gaps 2;

Db 744 atgcagataaagtgtgtaactgttcgaaaacagtgtggtggccaatttaacagagt 803
 Db 6304 ATATCAGATTAAATGTGGGATCTGTTGAGAACAGATGTGTTGATGAGTTCAACAGAGT 6363

Qy 804 tgcggttcccaagaagaatgtgtcccggaatccgagtgtggtgaattcccggttcc 863
 Db 6364 TGCCTGTGCGAAGAAAAAGTGTGTTCTTGAAGAACTCATCGGAGAAATTCCTGCCGCC 6423
 Qy 864 ggaatgtatgcagtggttcaaatlttaacatgaagaacttagtgggaagtgtatat 923
 Db 6424 AGACCTTGTGTTCTTGTGACAGAACTTCACATCGGACTTTAACGGAGAGTGTATCAT 6483
 Qy 924 aacaaatgttlaaatccatattgatgcatlttgatgtcaacttcaatgaatttcaat 983
 Db 6484 TCAAGTGGCTTGAATTCACACTTTGATGCTTGACTGCGACGTCGATGATTCACAC 6543
 Qy 984 ggaagaatg---ataactgttgggaacttaacatggcgaataaacttggatgtgtg 1040
 Db 6544 AAGAGGTGACAAAGCTTGTGGAAACATCTTGGAGATTAAGACCTTAGACAGTGG 6603
 Qy 1041 ttcttlaactgcagtcgtgtgcaaacatttgttcaagatccagatcttccctggaact 1100
 Db 6604 ATTCTTTACTAGGTGACGCGCTACAAAATTTGTGCAAGATCTTACCAACTGCTGTCT 6663
 Qy 1101 ttataatcatgaacaatgaatttcttcaactcaagaatgaact----- 1142
 Db 6664 CTCAATCATGACAAAGATCACTTCTACTATCAAGATGACTGTATTAATCATCTGTATC 6723
 Qy 1142 ----- 1142
 Db 6724 CACCACAGAGGAGAGTGTGATAGTGTGCTATTTCTTTGGCTTACCAAGCTTTCAT 6783
 Qy 1142 -----gttacaattatcttcccaatccgaatacgaacaacccgatatatacat 1189
 Db 6784 ATGATTTTTCGAGATATATCTCTGTCATCAAGATAGAGATTAACCTTAACACTATATAT 6843
 Qy 1190 tctataactacagagtcgaacagcgaatggatggatggatggatggatggatggatggat 1249
 Db 6844 TTGTATACTACGCTGGCGCAAAAGATCTTGGAGATGATAGTGGTGGATGATGATGATGAT 6903
 Qy 1250 cccgaagcccgacactcccggaatcgatcatcccaactcaaaacgaagcgaacatccg 1309
 Db 6904 CGAGAGTTCCTTATTACCAATAGCATTTATACAGAACTGGAAGACGCAAAAAAGCA 6963
 Qy 1310 tgggtcgagacttcaacatttcaacacacacacacacacacacacacacacacacacac 1369
 Db 6964 TAGGCAAGACTTGCACATTCATTGAACGGATTAACACATGCTGCTGCAACTGCGG 7023
 Qy 1370 tgggtgaagccttgaagaacacagcgaagcgaagcgaagcgaagcgaagcgaagcgaag 1429
 Db 7024 TCGTGAGAGATTTGAGAGACAGTGGAGAGAGTGAAGAGATTAATCGTAAAGAGGTTG 7083
 Qy 1430 tagagatagaagaagagtggaagaagagtggaagaagtggaagaagtggaagaagtggaaga 1489
 Db 7084 AAGAGATTAAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7143
 Qy 1490 tgggtcgaagtgcttgaagagtggaagagtggaagagtggaagagtggaagagtggaagag 1549
 Db 7144 TGTTCGAGAGATTTGCGTGAAGATTTAATGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 7203
 Qy 1550 gggagtggaagtaagaagaagaagaagtaagtaagtaagtaagtaagtaagtaagtaag 1609
 Db 7204 GAGAGTTAAGTAAGAAG 7263
 Qy 1610 ttgaagaagccttggcgcgcttaccgatataggaagaacttagataaa 1657
 Db 7264 TTGAAGAAATTTGTTGGAGAAAGCTTGGCAATCAGAGAGAGTCAAGGTAGA 7311

RESULT 5
 ATU39452/c 2884 bp DNA PLN 10-FEB-1996
 LOCUS Arabidopsis thaliana syntaxin-related knolle mRNA, complete cds.
 DEFINITION U39452
 ACCESSION U39452
 NID g1184166
 VERSION U39452.1 GI:1184166

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
gene
CDS

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (bases 1 to 2596)
Lukowitz, M., Mayer, U. and Jurgens, G.
Cytochromes in the Arabidopsis embryo involves the syntaxin-related
KNOX1 gene product
Cell 84 (1), 61-71 (1996)
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (26-OCT-1995) Wolfgang Lukowitz, Lehrstuhl fuer
Entwicklungs-genetik, Universitaet Tuebingen, Spemannstrasse 37-39,
D-72076 Tuebingen, Germany
Location/Qualifiers
1. 2884
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
520. 1452
520. 1452
/gene="knox1"
/note="syntaxin-related"
/codon_start=1
/protein_id="AAC49163.1"
/db_xref="PDB:1I84167"
/db_xref="GI:1184167"
Translation: "MNDLMTKFSMSYVDLKKAKNDNEAGPDLKASTKADKMDN
LSTPLFEEYKAEKGLSELTARLEOYHEESKGVHAEKSKLRSNLSVETSGRK
AKSISKLEPMKAKKEIKRLSGTPVRSPTAVNGLRKLEKMEFOLRQKMS
KEVERKRTVYTGSHANDMEKIIIDNAGEELFRLAIOEHGKVLVEIDR
YDAAEIKESLLEHGVFLDMANVNSGSEMDIEHVNASHYVDGANEKLTAKS
HQRNSRKICITLIVLLILIVLPIITSSSS"
BASE COUNT 833 a 569 c 582 g 900 t
ORIGIN

Query Match 20.4%; Score 404.8; DB 8; Length 2884;
Best Local Similarity 66.2%; Pred. No. 3.3e-74;
Matches 667; Conservative 0; Mismatches 247; Indels 94; Gaps 2;

DB 744 atgcagataaatagtgtgactgttcgaaacagtggtggtgaccacattcaagagtg 803
DB 2849 ATATCAGATTAAATGCTGGGATCTGTTGAGAACAGTGTGATGAGTCAACGATG 2790
QY 804 tgcgggtccggaagaatgtgtgccccggaatcgatgtggaattcccggtcc 863
DB 2789 TGCTGTGCGAGCAAAAGTGTGCTTGAAGAAATCTGATCTCGAGAAATTTCTCCCC 2730
QY 864 ggtatgtaatgagtgatcaaaatttaaatgaagacttagtgggaagtgtatat 923
DB 2729 AAGACCTTGTCTTCTTACAGAACTTCAACATCTCGACTTTAACGGGAAGGTATCAT 2670
QY 924 aacaagtggttaactcaatttgatgcatcttgatgtgaacttaattcaattat 983
DB 2669 TACAAGTCTGAAATCAACCTTGTGATGCTTCCAGCTGCGATGATGATCCACAC 2610
QY 984 ggaagaatg---ataactgttggaacttaacatgagcgaataaanaacttgatgtg 1040
DB 2609 AGAAGGTGACACAACTTGTGGAACATCTCTTGAGAAATTAAGACCTTGACAGTG 2550
QY 1041 ttcttactgcatctgtgtgcaaacattgttcaagatccaggtcttccctgagcaat 1100
DB 2549 ATTCTTACTAGCTCAGCCGTACAAAAATTTGTGCAAGATCTTAACCAACCTGTGTCT 2490
QY 1101 ttataatcatgacaatgagttcttactacaaagatgact----- 1142

DB 2489 CTACATCATCCTCAACAGTACTCTTCACTATCAAGATAGTGTATACATCTTGTATC 2430
QY 1142 ----- 1142
DB 2429 CACCACGAGGAGAGCTTGTATGATGCTGTATTCTTTGGCCCTACACAGCTTTCAT 2370
QY 1142 -----gtacatatatcttcccaatcgaaacacccagatgaatcaat 1189
DB 2369 ATGATTTTGCAGATATATCCCTCATCAAGATAGAAATTAACCTGAAGACTATATAT 2310
QY 1190 tcttactacagaggtgcaacgcgcgatggatgagatagagtgaggtccgtgactaca 1249
DB 2309 TTGTATATACCTGCGGCAAAAGATGCTTGGATGATATGCTGCTGATGATATACA 2250
QY 1250 cccgagcccgacaccccgatcgatcccaactcaacaaagacagcccaatccg 1309
DB 2249 CGAAGATTTGTATTATCCCAATAGATATATACCAAGATCGAAATCGAAACACCAAAACCA 2190
QY 1310 tgggtcagacttaacaaatttcaacaacgcgaatagtgtggtggtgagctccat 1369
DB 2189 TAGCAGAGACTTCAGCACATTCATTAGAACGATATACATGTGCTCTGAACTGCGC 2130
QY 1370 tgggtgaaaggttggaaagaacagcgaaagagcgagagtgatgtgataaagaagctg 1429
DB 2129 TCGTGAGAGATTTGAGAACACACTGGAAGAGTGAAGAGATATCTTAAAGAGGTG 2070
QY 1430 tagagataagaagaggttgaanaaagaggtggaagaggtgagatgactgagatgact 1489
DB 2069 AAGNATGAGAGAGAGTGAAGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 2010
QY 1490 tgttcaagaggttcttgaagaggttgaagaggttgcacacagaatgaagaattgtga 1549
DB 2009 TCTTCAGAGATGCTTCAAGATTTATGATGATGATGATGATGATGATGATGATGATG 1950
QY 1550 gggaggttgaagaagaagaagaagaatctggaatggaactcaaatggaagcagcagag 1609
DB 1949 GAGAGTAGAGTGAAGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 1890
QY 1610 tgaagaagcttcttggcgcggttaccgactgaagaacttagataaa 1657
DB 1889 TTGAAATGTTGTTGGAAAGCTTTCATCAGAGAGTGAAGAGTGAAGAGTGAAGAGTGA 1842

RESULT 6
AC006932 89493 bp DNA PLN 20-MAY-1999
LOCUS Genomic sequence for Arabidopsis thaliana BAC T2767 from chromosome
DEFINITION I, complete sequence.
ACCESSION AC006932
NID 94878038
VERSION AC006932.5 GI:4878038
KEYWORDS HMG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (bases 1 to 89493)
Walker, M., Shim, P., Brooks, S., Buehler, E., Chao, Q., Dunn, P.,
Khan, S., Kim, C., Altieri, H., Araujo, R., Conn, L., Conway, A.B.,
Gonzalez, J., Hansen, N.F., Huizar, L., Kremenetskaia, I., Lenz, C.,
Li, J., Liu, S., Lueros, S., Rowley, D., Schwartz, J., Toriumi, M.,
Vysotskaya, V., Yu, G., Davis, R.W., Federspiel, N.A., Theologis, A. and
Ecker, J.R.
TITLE Genomic sequence for Arabidopsis thaliana BAC T2767 from chromosome
JOURNAL I
REFERENCE Unpublished (1999)
AUTHORS 2 (bases 1 to 89493)
TITLE Ecker, J.R.
JOURNAL Direct Submission
Submitted (02-MAR-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and

	repeat_region	/note="MLTLE repeat: matches 308. .449 of consensus" 66836. .67002 /note="MLTID repeat: matches 327. .502 of consensus"
	repeat_region	57084. .67479 /note="MLT2_internal repeat: matches 4965. .4544 of consensus"
	repeat_region	57687. .67897 /note="MLT2_internal repeat: matches 4344. .4139 of consensus"
	repeat_region	53016. .68694 /note="MLT2_internal repeat: matches 2856. .2119 of consensus"
	repeat_region	64694. .68945 /note="MLTLE repeat: matches 564. .307 of consensus"
	repeat_region	68984. .69160 /note="MLTID repeat: matches 311. .140 of consensus"
Query Match	2.6%; Score 51.8; DB 9; Length 118595; Best Local Similarity 50.6%; Pred. No. 0.14;	
Matches 125; Conservative	0; Mismatches 122; Indels	0; Gaps
Db 1371 ggtgggaagccttggaacacagcggaagggggagaaatttgtatagaagccgt	1430	
Db 48468 GGAGGAGGAGG:3GAAGAAGCAGCATGCAGACAGCATGTGAGACAGACAGA	48409	
QY 1431 agagataagaaagagagcttgaaaaagtgtgaaggttaagataactgcagatgccctt	1490	
Db 48408 AGAGGGGAG	48349	
QY 1491 gtttcagaagctgttccttgaagggtttaagagcttcacaagaatgaagaaatttgtag	1550	
Db 48348 AGAGAGGAGAGAGAGAGAGAGATGAGAGAAAGAGAGAGAGAGAGAGATGAGCA	48289	
QY 1551 ggagcttgatlaagaaaggaaggaatcttgatgaactcaattggaagcgactggaagt	1610	
Db 48288 GGAAGAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAGCA	48229	
QY 1611 tgaagaag_1617		
Db 48228 GGAAGAG 4822?		
RESULT 10		
LOCUS HSAJ9612		
DEFINITION HSAJ9612 153053 bp DNA HTG 11-NOV-1998		
SEQUENCE Homo sapiens chromosome 17 clone PAC 1577 map p11, WORKING DRAFT		
ACCESSION AJ009612		
NID 93581972		
VERSION AJ009612.1 GI:3581972		
KEYWORDS HTG; HTGS_PHASE1.		
SOURCE human.		
ORGANISM Homo sapiens		
REFERENCE Eukaryota: Metazoa: Chordata: Vertebrata: Mammalia: Eutheria:		
AUTHORS Primates: Catarrhini: Hominoidea: Homo.		
TITLE 1 (bases 1 to 153053)		
JOURNAL Radeloff,U., Hennig,S., Ramser,J., Francis,F., Steffens,C., Klein,M., Sezanski,P., Poustka,A., Reinhard,R. and Lehrach,H. Direct Submissio Submitted (17-JUL-1998) Steffen Hennig, MPIHG, Abt. Leirach, Max Planck Institut fuer Molekulare Genetik, Innestrasse 73, Berlin, 14195, Germany		
COMMENT HTGS phased: Unordered fragments separated by 800 N nucleotides. * NOTE: This is a working draft sequence. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.		
FEATURES location/Qualifiers		
source 1..153053		
organism="Homo sapiens"		
db_xref="taxon:9606"		
chromosome="17"		
clone="PAC 1577"		

RESULT	14				
AC005140/c					
LOCUS					
DEFINITION	AC005140	194410 bp	DNA	HTG	01-APR-1999
	Plasmodium falciparum	chromosome 12,	WORKING DRAFT SEQUENCE,	14	
ACCSSION	AC005140				
NID	94558572				
VERSION	AC005140..3	GI:4558572			
KEYWORDS	HTG; HTGS.PHASE1.				
SOURCE	malaria parasite P. falciparum.				
ORGANISM	Plasmodium falciparum				
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
AUTHORS	1 (bases 1 to 194410)				
TITLE	Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kuri,O.B. and Davis,R.W.				
REFERENCE	Plasmodium falciparum 3D7 chromosome 12				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 194410)				
AUTHORS	Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-JUN-1998) Stanford DNA Sequencing and Technology				
	Center, Stanford University, 855 California Avenue, Palo Alto, CA				
	94304, USA				
COMMENT	On Apr 2, 1999 this sequence version replaced v1.422165				

[illegible]

Query Match	2.5%	Score 49.2	DB 35	Length 194410
Host Local Similarity	52.4%	Pred. No. 0.46		
Matches 108	Conservative	0	Mismatches 98	Indels 0
Gaps 0				
QY	1651	agataaattcgtatgcacacatatatgcattgcattgattatgact	1710	

[illegible]

RESULT	15
AC006279/c	
LOCUS	
DEFINITION	
AC006279	219200 bp DNA
Plasmodium falciparum	chromosome 12, WORKING DRAFT SEQUENCE,
unordered pieces.	
	01-APR-1999

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (bases 1 to 21,920)	Hyman, R.W., Fung, E.L.,	plasmidium falcatum 3D7 chromosome 12	Unpublished	2 (bases 1 to 21920)
	Qin, F., Tamaki, T.,			
	Kurdi, O.B. and Davis, R.W.			

AUTHORS Hyman, R.W., Tang, A.L., & Kornberg, J.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
Stanford University, 855 California Avenue, Palo Alto, CA

On April 2, 2019, this sequence version replaced gi:14337165.
NOTE: This is a 'working draft' sequence. It currently
consists of 6 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved. 77526 contigs = 77526 bp in length

77526	contig of 77526 bp in length
77527	gap of unknown length
77722	gap of 52405 bp in length
140131	gap of unknown length
140132	gap of unknown length
140332	contig of 33177 bp in length
133708	gap of unknown length
133709	gap of 39089 bp in length
133710	contig of unknown length
212997	gap of unknown length
212998	gap of 1907 bp in length
212999	contig of unknown length
213004	gap of unknown length
219500	contig of 4096 bp in length
219501	contig of unknown length

FEATURES	Location/Qualifiers
source	1. . 219200
	/cc=us="Plasmodium falciparum"

BASE COUNT	ORIGIN	88210 a	22418 c	19532 g	88036 t	1004 others

Query Match	2.5%	Score 49.2	DB 35	Length 219200
Best Local Similarity	52.4%	Pred. No. 0.45		
Matches 108	Conservative	0	Mismatches 98	Indels 0
			Gaps	0

DQ 1651 agataattcgaatgatgcatcgacacatalatatgctacatcgttaaaactct 1710
 + + + + + | + + + + + | + + + + + | + + + + + |
 Db 111422 AAAAAAAAAATTGATATTAATATATATATATATATATAATTTTAAATAAATAAGTTAAAT 111363

•

Thu Oct 14 07:57:23 1999

us-09-075-375-1.rng

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 1999, 00:56:50 : Search time 183.21 Seconds
(without alignments)
2705.259 Million cell updates/sec

Title: us-09-075-375-1
Perfect score: 1881
Sequence: 1 tgggttgctgaatttacc.....aattttgctaaaaaaa 1981

Scoring table: IDENTITY_NUC

Searched: 311585 segs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1894.6	95.6	1981	1	T66241	Romaine lettuce vi
2	643.8	32.5	1589	1	T66242	Tobacco violaxanth
3	610.2	30.8	1555	1	T66243	Arabidopsis violax
4	44.6	2.3	425	1	X41515	Human secreted pro
5	42.4	2.1	3901	1	T89345	Human p160 CDNA 16
6	42.4	2.1	3211	1	T89345	Human p160 CDNA 16
7	42.2	2.1	15464	1	Q11415	Ryanodine receptor
8	41.8	2.1	110000	1	X20248_02	Continuation (3 of
9	41.6	2.1	507	1	Q32658	LEKMI 10 CDNA. New
10	41.2	2.1	1505	1	O55750	Genomic clone G11F
11	41.2	2.1	9636	1	O67190	P. falciparum tran
12	41.1	2.1	110000	1	V21209_00	Methanococcus jann
13	40.8	2.1	1484	1	V74906	Staphylococcus aur
14	39.6	2.0	7215	1	X13039	Enterococcus faeca
15	39.6	2.0	591	1	X20389	Borrelia burgdorfe
16	39.4	2.0	9399	1	X20261	Borrelia burgdorfe
17	38.2	2.0	110000	1	V21209_03	Continuation (4 of
18	38.8	2.0	3088	1	Q13727	Plasmidium falcipa
19	38.6	1.9	6942	1	O86480	Merosin partial CD
20	38.6	1.9	9534	1	T17419	Metrosin CDNA. New
21	38.6	1.9	53585	1	X20251	Borrelia burgdorfe
22	38.4	1.9	562	1	V64661	Seq ID 29 from DE1
23	38.4	1.9	26811	1	X20253	Borrelia burgdorfe
24	38.2	1.9	9542	1	X20260	Borrelia burgdorfe
25	38	1.9	4590	1	N60472	Sequence encoding
26	38	1.9	9048	1	T43225	Brassica napus FCA
27	38	1.9	110000	1	V21209_00	Methanococcus jann
28	37.8	1.9	2643	1	Q39212	CENP-B CDNA. Human
29	37.8	1.9	3891	1	T28244	C. botulinum type
30	37.8	1.9	1546	1	V30575	Clostridium botuli
31	37.6	1.9	13508	1	V74431	Staphylococcus aur
32	37.4	1.9	15377	1	Q25975	MM mutant porcine
33	37.4	1.9	6106	1	Q74678	Bacillus cereus ve
34	37.4	1.9	6049	1	T13939	B. cereus VIP2(a)
35	37.4	1.9	3224	1	T91872	Rat V1521 gene. Br
36	37.4	1.9	6049	1	T73993	B. cereus VIP2(a)
37	37.2	1.9	997	1	Q22702	DNA encoding veget
38	37.2	1.9	9789	1	T41852	Sequence encoding
39	37.2	1.9	29392	1	V14122	CDNA encoding plas
40	37.2	1.9	9542	1	X20260	Mouse poly Ig rece
41	37.2	1.9	53577	1	T18551	Borrelia burgdorfe
42	37	1.9	6152	1	T78867	Human polyomelic k
43	37	1.9	6152	1	T78867	P. falciparum live

44 37 1.9 5361 1 T78868
45 37 1.9 4982 1 T85267

ALIGNMENTS

```
RESULT 1
ID T66241 standard: CDNA: 1981 BP.
AC T66241:
DN 28-JUL-1997 (first entry)
DE Romaine lettuce violaxanthin de-epoxidase CDNA.
KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KM Photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW xanthophyll; let'nice; ss.
OS Lactuca sativa L. cv. romaine.
FH Key location/Qualifiers
FT misc_difference 26..29
FT /tag= a
FT /note= "bases 26-29 are illegible in Fig 1"
FT misc_difference 66..72
FT /tag= b
FT /note= "bases 66-72 are illegible in Fig 1"
FT misc_difference 105..110
FT /tag= c
FT /note= "bases 105-110 are illegible in Fig 1"
FT misc_difference 147..149
FT /tag= d
FT /note= "bases 147-149 are illegible in Fig 1"
FT misc_difference 186..189
FT /tag= e
FT /note= "bases 186-189 are illegible in Fig 1"
FT misc_difference 226..227
FT /tag= f
FT /note= "bases 226-227 are illegible in Fig 1"
FT cds
FT /tag= g
FT /note= "bases 226-227 are illegible in Fig 1"
FT transp-epoxide 235..609
FT /tag= h
FT /note= "bases 235-609 are illegible in Fig 1"
FT mat-peptide 610..1653
FT /tag= i
FT /note= "bases 610-1653 are illegible in Fig 1"
PI WO9717447-A2.
PD 15-MAY-1997.
PF 07-NOV-1996: US-006315.
PR 06-AUG-1996: US-023502.
PA (CALU) CALGENE INC. Yamamoto HY;
PI Bugos RC, Rockholm DC,
DR WPI: 97-281036/23.
PI P-PSDB: W09874.
PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
PT sensitivity of a plant to light
PS Claim 3, Fig 1: 4pp: English.
CC A CDNA clone (T66241) codes for the 55 kDa violaxanthin de-epoxidase
CC (VDE) (WO9874) of romaine lettuce. VDE was purified from romaine
CC lettuce chloroplasts and 2 tryptic peptides were used to develop
CC primers (see also T66244-45), which amplified a partial VDE
CC sequence. The amplified sequence was then used to screen a lettuce
CC cDNA library, and the 1981 bp DNA sequence was identified. VDE
CC orientation, can be used in genetic constructs, pref. also contg. a
CC plastid translocation sequence, to modify VDE levels in plants.
CC Increased levels result in the plant being tolerant of increased
CC light and therefore more productive and/or more resistant to
CC disease. Underexpression of VDE increases photosynthetic
CC efficiency under low light. The photosensitivity of a range of
CC crops, trees and ornamentals can be modified.
CC Sequence 1981 BP: 608 A: 337 C: 433 G: 577 T:
SO
```

Query Match 95.6%: Score 1894.6; DB 1: Length 1981;
Best Local Similarity 97.4%: Pred. No. 0;

Matches 1931:

Conservative

0;

Mismatches

49;

Indels

2;

Gaps

2;

QY 1 tctggttcgaatttcccccacacgaatttctccacacaaatgagataagagct 60
 Db 1 TGTGGTTGATTTTACCCACACNNNTTTCCTACCATTAATGGATAGAGCT 60
 QY 61 aattcccttgacaatttcccaatttctccctccgacacacataataatgtagc 120
 Db 61 AATTNNNNNNNNCAATTTTCCAAATTTCTTCCCTCCGCCACACTANNNNNNACTGACG 120
 QY 121 cactcgaacgtacacatcttggaaaaaagcgcgaatttccaaagaacgaagaata 180
 Db 121 CACTTGAAGCTACATGTTTGAANNNGAGTAGATTTACAAGAAGCAGAGATTA 180
 QY 181 taagc-ttcaagtaactcgcagtcagtgagtgagccttggaaacgaacaaatgtagc 239
 Db 181 TAAGCANNNAAGTACTCCGATGTGAGTGCGCTTGSAMGCC-ACNNACTGGCTATGCG 239
 QY 240 tcttctcctcaactgatttctctgcaagaagaagccctcaattatataagagtc 299
 Db 240 TCTTCCCTTCAACCGTATTTCTCTGCAAGAAGAGAGCCCTCAATTATATGCAAGATC 299
 QY 300 accatgtaagtaaggttccaaagagtgagcaacccctcacaacataatcatgtaga 359
 Db 300 ACCATGTATGATAAAGGTTTCAAGAGAGGACAACTCTTACCAACATATATATGATGA 359
 QY 360 aattgcacacaatgatatatttaattcttccggattgtttacatacctaagaaga 419
 Db 360 AATTGATCTTAACATGATATTTTATTTCTTCCGTTGTTTACATCTTATTAACACAG 419
 QY 420 tcttctcgaattctgacatgcaagaataatctcagatatacagatcagataag 479
 Db 420 TCTTCTTCAATTTTACGATTTGCAAGATTAATCTCAGATATGCAACATTTGATACAG 479
 QY 480 tttcgaagaatacaagaattgatactcaaaaggagcagatgtagcttccctgaaga 539
 Db 480 TTTTGAAGAATACAAAGATTGATTCATAAAGGGSCATGACTTGTGATTTGTAAGCA 539
 QY 540 atgagacaattcacaatggtcgtatgtagtctgacatttcttccgag 589
 Db 540 ATGAGACATTTCAATGATTTGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 589
 QY 600 agtgcgtcgttgaatgctcctaaactgctgcttcttactcaagaatgtagcag 599
 Db 600 AGTGTGATGCGTGTGATGCTTAAACCTTGTGCTTGTACTCAAAAGAAAGCAGATGA 599
 QY 660 gcttgcacaaatgataagaacccaactctgctgcgcaaacgctgctgctcagactg 719
 Db 660 GCTTGCACAAATGTATAGCAAAACCCATCTTGTGCGGCAAAAGCTGCTGCTACAGACTG 719
 QY 720 caacaatcgtcctgacagacgaatgctcagataaattggtgactgttccgaagaag 779
 Db 720 CAACAATGCTCTGACAGACCGAATCTCAGATTAATATGTTGGTGTGTAAGAAACAG 779
 QY 780 tctgtagacacaaatcagaagtgctgcttccgaagaagaattggtcccggaatc 839
 Db 780 TGTGTGAGACCAATTCACAGAGTGTGCTGCTTCCGGAAGAATGTGCTCCCGGAATC 839
 QY 840 ggaatggtggaatctcccggttccgagatcgatgaagtggttcaaaatcttaacatga 899
 Db 840 GGAATGTTGGGATTCCTCCGTTCCGATGTGATGAGTGTGTAAGAAATTTTAAACATGA 899
 QY 900 agacttaagtggaagtggtatatacaagtggttcaatcctcattatgtagatga 959
 Db 900 AGACTTTAGTGGAGTGTATATTAACAAGTGTTAATCCACATTTGATGATTTGA 959
 QY 960 ttgtaactcaatggttcatatagaaaatgataactgttgggaacttaacatgagc 1019
 Db 960 TTGTCACTTCAATGAGTTTCAATGAAATGATGAATCTTGTGGAACTTAAACATGAGCG 1019
 QY 1020 cataaaacttggatggttcttactcagatcgtcgtgcaacaatttgcataaag 1079
 Db 1020 CATAAAACTTTGATGTGTGTTTACTCGATGCTGTGCAAAACATTTGTTCAAGA 1079

QY 1080 tccagatcttctgagacacttataatcatgacaatgagttcttccacgaatga 1139
 Db 1080 TCCAGATCTTCTGAGACACTTATATCATGATGATGTTTCTTCACTACCAAGATGA 1139
 QY 1140 ctggtacataatcttcccaatcgaagaaccccgatgattacataltgatacta 1199
 Db 1140 CTGTTACATATATCTTCCCAATTCGAACCAACCCGACATTAACATGATGATCTA 1199
 QY 1200 ccgaggtcgaagaagcagatgagatgagatgagtgagtcggtgatacccggaagcc 1259
 Db 1200 CCGAGTTCGAAGCAGCATGAGATGAGTACGGTGGCTGTATTAACCCGAAGTCC 1259
 QY 1260 gaaactcccgatcgatcatcccaactcgaagaagaagcgaatcgttggtcaga 1319
 Db 1260 GAACTCCCGAATCGACATCCCAACTCAAAAAGCAGCAAAATCCGTGGTGCAGAA 1319
 QY 1320 cttaacaattcacaacacgaacaaatagttgtggcctgagcctcattgtgaaag 1379
 Db 1320 CTTTAAACATTTTCAATACACCAATAGTGTGGCTGAGCTTCAATGATGGAAG 1379
 QY 1380 gcttgaagaacagcgaagaagcgaagaagtggtgataaagaagcgttagagata 1439
 Db 1380 GCTTGAAGAAGCAGCGAAGCGGAGAGGTTGATTAAGAAGCTGTAGATAGA 1439
 QY 1440 aagaaggttgaagaagaagtgagaagtgtagatactgagatgactgtgtcagag 1499
 Db 1440 AGAAGGTTTAAAGAGAGTGAAGAGTTGAGATCACTGATGATGATGATGATGATG 1499
 QY 1500 gttccttgaaggttgaagaagtgtagaagatgaagaatgttgtgaggaagtgag 1559
 Db 1500 GTTCTTGAAGGTTTAAAGAGTTGCAACAGATTAAGAAGATTTTGAAGGAGTGAAG 1559
 QY 1560 taaagaagaagaagaatctgaatgaactcaaatggaagcgaactgaatgaagaagc 1619
 Db 1560 TAAAGAAGAGAGAGATTTCTGAATGAACCTCAATGGAAGCAGTGAATGGAAGCT 1619
 QY 1620 ttttggcgcggttaccgagatgaagaacttagataaatttgatgattcagacaa 1679
 Db 1620 TTTTGGCGCGGTACCGATGAGAACTTGAATTAATTTGATGATGATGATGATGATG 1679
 QY 1680 tatataatcctatgagattttagatatactagaagaacccaacaaacttggatag 1739
 Db 1680 TATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1739
 QY 1740 gtagataagcgttggatgttattttagcttaaaatgtagaagcttcttaatt 1799
 Db 1740 GTGATTAAGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1799
 QY 1800 cttaacaaaattgatttctatgtagtgaagaagttcagaagaagcgtgaagaag 1859
 Db 1800 CTTTACAAAATTTGATTTGCTCATTGTGAGCCAAAGGTTTACAAAAGCTGAAGAG 1859
 QY 1860 tcttgcggttgcggttgaagcgaatcttcttggcggaagtggttgcagctg 1919
 Db 1860 CTTGCGGTTTCCGTTGAGGCAAAATTTTGGCGGCAATGCGTGTGACCGGCTTT 1919
 QY 1920 tctttaaacaatgatttctttaaataatgtagatcttcaatttcttgcataaaga 1979
 Db 1920 TCTTTTAAACATGATTTTATTAATGATGATGATTTTCAATTTTGTGCTAAAGAA 1979
 QY 1980 aa 1981
 Db 1980 AA 1981

RESULT 2

T66242

ID T66242 standard; cDNA; 1589 BP.

AC T66242;

DT 28-JUL-1997 (first entry)

DE Tobacco violaxanthin de-epoxidase cDNA.

KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;

KM photoprotection: transgenic plant; zeaxanthin; antheraxanthin;
 KM xanthophyll; lettuce: ss.
 OS Nicotiana tabacum L. cv. xanthi.
 FH Key Location/Qualifiers
 FT cds 42..1478
 FT mat_peptide 42..443
 FT /tag= a
 FT /tag= b
 FT /tag= 444..1475
 FT /tag= c
 PN MO9717447-A2.
 PD 15-MAY-1997.
 PE 07-NOV-1996; US-018291.
 PR 07-NOV-1996; US-006315.
 PR 06-AUG-1996; US-023502.
 PA (CAL) CALGENE INC.
 PI Bugos RC, Rockholm DC, Yamamoto HY;
 PI MPI: 97-281036/25.
 DR P-PSDB: M09875.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light.
 PS Sensitivity: Fig 2, 41pp: English.
 PS Disclosure: Fig 2, 41pp: English.
 CC A cDNA clone (T66242) codes for tobacco violaxanthin de-epoxidase
 CC (VDE) (M09875), an enzyme that catalyses the de-epoxidation of
 CC violaxanthin to zeaxanthin and antheraxanthin. VDE nucleic acids
 CC (see also T66241, T66243), in sense or antisense orientation, can
 CC be used in genetic constructs, pref. also contg. a plastid
 CC translocation sequence, to modify VDE levels in plants. Increased
 CC levels result in the plant being tolerant of increased light and
 CC therefore more productive and/or more resistant to disease.
 CC Underexpression of VDE increases photosynthetic efficiency under
 CC low light. The photosensitivity of a range of crops, trees and
 CC ornamentals can be modified.
 SO Sequence 1589 BP: 500 A: 261 C: 367 G: 461 T:

Query Match 32.5%; Score 643.8; DB 1: Length 1589;
 Best Local Similarity 71.8%; Pred. No. 2.2e-154;
 Matches 862; Conservative 0; Mismatches 327; Indels 12; Gaps 1;

QY 483 tgaggaataacaaagatttgctcaaaagggcatgacttcttcttgaagaacatg 542
 DB 317 TAAGGGAATTCAGATGCCAATTTCCAAAGATGTAGCTTGATGGTTGGAGAAAGT 376
 QY 543 gagaacattacatacaatttgctatcgatgttgcacattgtttgtcccgagatg 602
 DB 377 GGGCCAAATTTGCCAAAGACGATGTAGCTATATTCATTTTGTCACTTCCTCAAAAGC 436
 QY 603 tgatgccttgatgtcttaaaacttgctgttactcaaaagatgagatgact 662
 DB 437 TGATGGGTTGATGCTCTCAAGACTTGTACTTGCTTACTGAAAGATGCGAGGTAGAGCT 496
 QY 663 tgcgaatgtatagcaaacacacttgctgctgcaaaagcttgcctgctcaagactgca 722
 DB 497 TCGGAGATGCTATTCGAAACCTCGCATGTGACATGATGTGCTGCTCCGAGACTTGCAA 556
 QY 723 caatgctctgagagacccgaatgtcgaataaagtgtgaactgttgcgaagaacatgt 782
 DB 557 CAATTAACCTCGAGAAAGGAAATGTAGATTAAGTGTGATTTGTTGTTGAAACAGAGCT 616
 QY 783 gctggaacattcaacagatgtgctgttcccgaaagaatgtgtgcccgggaatcgca 842
 DB 617 CGTAGAGAGATTCATAGAGTGTGCACTCCGGAAGAAATGTATACCTCGTAATCTGA 676
 QY 843 tctgtgtgaattcccggttccgcatgtatgtgcaagtggtgtcaaaatttaacatgaaga 902
 DB 677 TGTGTGTGATCTTCCTTACCTCATCCAGTGTCTTGTGCTCCGAGACTTGAATGAAGA 736
 QY 903 cttaagtgggaagtgtgatatcaacagtggtttaaattcaattgtatgtcatgtgcatg 962
 DB 737 TTTTACGGGGAATGTGTTCACTTACTCGGGTTGCAATCCACTTTTGATGCTTTGATG 796
 QY 963 tcaacttcatgattcatatgaaaaatgataaactgttgggaacttaacatgtgagcat 1022

DB 797 CCAATTCGATGAGTTCATACAGAGAAACAAACTGTGGGAATTTATCTTGAGAGAT 856
 QY 1023 aaaaacttggaagggtgttcttaactcgatctgtgcaaacattgttcaagatcc 1082
 DB 857 ACGTACACCTCATGAGAGATTTTACTCGATCGGGGTCAGAAATTTGTGCAGATCC 916
 QY 1083 agatcttccttgagacatttaataatgacatgacatgagatttcaactcaacagatgact 1142
 DB 917 AAGTATCCGGAGATCTCTACATCTATGATATGATATCTCTCTACCAAGATGACTG 976
 QY 1143 gtaacattatcttcccaaaacgcgaacaaacccgatgatacatattcgtatcacccg 1202
 DB 977 GTATATTTTGTGATCCAGAAAGTAAATAGCCAGAGATTTACTATTTGTGATTA 1036
 QY 1203 aggtcgaaacgctgcattgagatgacgtgtgggtccgctgacatcaacccgaagccgac 1262
 DB 1037 GGGCAGAAATGATGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1096
 QY 1263 actcccgatcgtatcatcccaaacctaacaaagcaagcaaatccgtgtgagatgact 1322
 DB 1097 TTTGCTGAAACATTATACCGAGATTCGAAACCGAGCTCAAAAGTTGGCGTGATTT 1156
 QY 1323 taacaattcatacaacacccgaacataatgtgtggtcctgagacatcattgtgtgaagct 1382
 DB 1157 CAACACATTCATTAATAAAGCAACATATCATATGATGATGATGATGATGATGATGAT 1216
 QY 1383 tgagaaacagcgcgaagagagcgagagatgtgtgttaaaagacgtgtgaagatagaaga 1442
 DB 1217 GGAAGAAAGTGAAG 1276
 QY 1443 agaggttgaanaagaggtgtgagaggttgaagatctgagatgactgtgttcaagaggt 1502
 DB 1277 AGAAGTA-----GAGAAAGTGAAGATTAAGAGATCACTTATTCAGTAACCT 1324
 QY 1503 gcttgaaggttgaagaggttgaagaggttgaagagatgtgtgtgagaggttgaagtaa 1562
 DB 1325 GTTGAAGGTTTGAAGAGTCCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1384
 QY 1563 agaaagaaagaaatttgaatgaacttcaaatggaagcgtgagatgtgaaagctttt 1622
 DB 1385 AGAAGAAATGAGATTTTGTGATGACTTAAATGGAAGCACTGAGATGAGAAACTTTT 1444
 QY 1623 tgggcgcgcttaccgatttggaacttgataaatttgatgattgattcagacaatat 1682
 DB 1445 TGGGCTGCTTTACCAATAAGAAATTAAAGTAAATTTTAAACTCAACATATAT 1504
 QY 1683 A 1683
 DB 1505 A 1505

RESULT 3
 ID T66243 standard; cDNA: 1555 BP.
 AC T66243.
 DT 28-JUL-1997 (first entry)
 DE Arabidopsis violaxanthin de-epoxidase cDNA.
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll; lettuce: ss.
 OS Arabidopsis thaliana cv. Columbia.
 FH Key Location/Qualifiers
 FT cds 45..1433
 FT mat_peptide 45..383
 FT /tag= a
 FT /tag= b
 FT /tag= 384..1430
 FT /tag= c
 PN MO9717447-A2.
 PD 15-MAY-1997.
 PE 07-NOV-1996; US-018291.
 PR 07-NOV-1996; US-006315.

PR 06-AUG-1996: US-023502.
 PA (CALJ) CALGENE INC.
 PI Bugos RC, Rockholm DC, Yamamoto HY.
 DR WPI: 97-281036/25.
 P-PSDB: W09876.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PS sensitivity of a plant to light
 CC A cDNA clone (T66243) codes for Arabidopsis violaxanthin
 CC de-epoxidase (VDE) (W09876), an enzyme that catalyses the
 CC VDE nucleic acids (see also T66241-42), in sense or antisense
 CC orientation, can be used in genetic constructs, pref. also contg. a
 CC plastid translocation sequence, to modify VDE levels in plants.
 CC Increased levels result in the plant being tolerant of increased
 CC light and therefore more productive and/or more resistant to
 CC disease. Underexpression of VDE increases photosynthetic
 CC efficiency under low light. The photosensitivity of a range of
 CC crops, trees and ornamentals can be modified.
 SQ Sequence 1555 BP; 485 A; 286 C; 375 G; 409 T;

Query Match 30.8%; Score 610.2; DB 1: Length 1555;
 Best Local Similarity 72.5%; Pred. No. 7, 6e-146;
 Matches 805; Conservative 0; Mismatches 303; Indels 3; Gaps 1;

QY 550 ttcaacaattgctatcgtatgttgcacattgttatcgtccagagttgatcc 609
 DB 324 TTGTTAAACTCGTGGGTGTTTAGCTTGCGCTTCTTATGTTCCATCTGCATATCA 383
 QY 610 gttgagtccttaaaactgtgttttaccacaagaatgagagttgagcaaa 659
 DB 384 GTTGATGCACTTAAACTGTGTCATCTTATGGAAGGATGAGATGCAATCGCAAG 443
 QY 670 tttatagcaaacccatctgttcgagcaacgttgcctgtctacagacttgaacaatgt 729
 DB 444 TGATTTCCCAACCTGCTGTCGAGCAATGTCGCTTCCCTTACAGCTTGAATTAACCGT 503
 QY 730 cctgacagacccaatgctacagataaagtgtgagctgttcgcaaaaagctgtgagac 789
 DB 504 CCAGATGAAGAACGAGTGCAGATTAATGTGGGATCTGTTTGAGAACAGGTGTTGAT 563
 QY 790 caattcaagagtggtggtgttcccgaaagaatgtgtcccggaatcgagatgtggt 849
 DB 564 GAGTTCAACGAGTGTGCTGTGCGAAGAAAGTGTGTTCTTAAGAAATCGATCTGGA 823
 QY 850 gaattcccggttcgagatcgaatgcagtggtgttcaaaatttaacatgaagacttga 909
 DB 624 GAATTTCCGCCCCAGACCTTCTGTTCTTACAGAACTTCAACATCTCGGACTTTAAC 683
 QY 910 gggagagtggtatataaagaatgtttaaactcacttgaatgagttgagttcaact 969
 DB 684 GGGAGGTGTGTCATTACAGTGGCTTAATCCACACTTGTATGCTTGCAGCTCCGAGT 743
 QY 970 catgagttcatatggaatgt--ataaactgttgggaacttaacatgagcacaanaa 1026
 DB 744 CATGAGTTCACACAGAAAGTGCACAAAGCTTGTGGAACAATCTTGGAGAAATTAAG 803
 QY 1027 acttgatgtgtgttcttactcgaatcgtgtgcaaacattgtttaaataatcagat 1086
 DB 804 ACCCTAAGACGTGATCTTACTAGGTGAGCGGTACAAAATTCGTGCAAAATCTTAAC 863
 QY 1087 ctctcggagcacttataatcatgacaatgagttcttccactcaacaagaatgagttac 1146
 DB 864 CAACCTGTTCTCTCAATCAATGACAAAGATACCTTCACTATCAAGATACATGAT 923
 QY 1147 atattatctccaaatcgaaacaaaccgagatgatatcatatttggttactaccagagt 1206
 DB 924 ATCTGTATCAACAAGATAGAAATTAACCTGAAGATATATTTGTACTACCGTGG 983
 QY 1207 cgaacagcagatgagatgagatgagtggtgtcgtgatatcacccgaacgagcagactc 1286
 DB 984 CGAAGCATGCTTGGCATGATATGCTGTGATGTAACAGAGAGAGTTCTGTATTA 1043

QY 1267 cccgagatcatcaccacccctacaaagacagccaaatccgtgtgagacttacc 1326
 DB 1044 CCCAATGACATTATTCAGAGACTGAAAGAACGACAAAAGCATAGCGAATCTTACG 1103
 QY 1327 aattcacaacaccgacaatagtgtggtcctgagcctccatctgtgtggaagcgttga 1386
 DB 1104 ACATTCTTTGACACGATTAACATGTGTGCTGAAACCTGCGCTGTGAGAGATTTAG 1163
 QY 1387 aaacacggaagaagggcgagaagttgttgataaagaagcgtgtgagatgaagaag 1446
 DB 1164 AAGCAGATGGAAGAGAGTGAAGATTAATCGTAAAGAGGTGAGACATTAAGAGAG 1223
 QY 1447 gtttaaaagagtggaagagttgagatgagatgagacttgtttcagaagttgct 1506
 DB 1224 GTAGAGAGAGAAATGAGAGAGGTGCTAGACATGAGATGACCTTGTTCAGAGATTGGCT 1283
 QY 1507 gaagggttaagagtgtaacaagaatgaagaatgtgtgtgagaggtgagtaagaag 1566
 DB 1284 GAAGGATTTAATGAAGTGAAGCAGAGAGATTTCTGACAGAGTTAAGTAAGAA 1343
 QY 1567 gagaagaaatctcgaatgaactcaaatggaagcagactgagttgaagcgttttgg 1626
 DB 1344 GAGATGAGATTTTGGATGAGATCAAAATGGAAGCAAGTGAAGTGAAGATTTGTTGG 1403
 QY 1627 cgcgcgttacccgattaggaactagataa 1657
 DB 1404 AAGCTTTGCCAATCAGAGAGTACGTAGA 1434

RESULT 4
 X41515
 ID X41515 standard; cDNA; 425 BP.
 AC X41515.
 DE 22-JUN-1999 (first entry)
 DT Human secreted protein 5' EST SPQ ID NO: 174 from WO 900553.
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW Homo sapiens.
 OS Homo sapiens.
 PN W0906553-A2.
 PD 11-FEB-1999.
 PF 31-JUL-1998: IB1237.
 PR 01-AUG-1997: US-905051.
 PA (GENST) GENSET.
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 DR WPI: 99-153783/13.
 DR P-PSDB: Y12657.
 PT New nucleic acids encoding human secreted proteins - obtained from
 PT cDNA libraries derived from umbilical cord, lymph ganglia,
 PT lymphocytes and placental tissue
 PS Claim 1, Page 306; 411p; English.
 CC X41379 to X41526 represent 5' expressed sequence tags (ESTs) for human
 CC secreted proteins, and encode the proteins given in Y12521 to Y12658,
 CC respectively. The proteins given represent the signal peptide and an
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences
 CC can be used for producing secreted human gene products. They can also
 CC be used to develop products for diagnosis and therapy. The proteins
 CC obtained may have cytokine activity, cell proliferation/differentiation
 CC activity, haemostatic activity, tissue growth regulating
 CC activity, reproductive hormone regulating activity, chemotactic/
 CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
 CC or other activities. The products can be used in forensic, gene therapy
 CC and chromosome mapping procedures. The sequences can also be used for
 CC obtaining corresponding promoter sequences. The nucleic acids encoding
 CC the signal peptide can be used for directing extracellular secretion of
 CC a polypeptide or the insertion of a polypeptide into a membrane, or
 CC importing a polypeptide into a cell.
 SQ Sequence 425 BP; 164 A; 63 C; 99 G; 95 T;

Query Match	2.38;	Score 44.6;	DB 1;	Length 425;
Best Local Similarity	44.9%;	Pred. No. 0.025;		
Matches 153; Conservative	4;	Mismatches 184;	Indels 0;	Gaps 0

[illegible]

RESULT 5
 ID T89345 standard: cDNA; 3901 BP.
 DT T89345:
 DE 11-MAR-1998 (first entry)
 KW Human p160 cDNA 160.1.
 KW p160; p62; cytoplasmic; T cell; B cell; development; activation;
 KW modulation; cellular response; cell proliferation; autoimmune disease;
 KW p56-lck; ss.
 KW Homo sapiens.
 OS
 FH Key location/Qualifiers
 FT CDS 439..3846
 FT /*tag= a
 FT /*product= p160
 FT /note= "160.1"
 PD MO9722255-A1.
 PD 26-JUN-1997.
 PF 11-DEC-1996: U19944.
 PR 19-DEC-1995: US-574959.
 PR (DAND) DANA FARBER CANCER INST INC.
 PA JOUNG I, Shin J, Strominger JL, Vadmudi RK;
 PI WPI: 97-341351/31.
 DR P-PSDB: W31185.
 PT cDNA encoding p62 and p160 and corresponding proteins - used in the
 PT treatment of autoimmune disease and for T and B cell proliferation,
 PT e.g. for treatment of tumours
 PS Claim 82: Fig 8; 175pp: English.
 PS This cDNA sequence encodes a novel p160 (160.1) which is capable
 CC of activating transcription of a variety of genes upon activation of p62
 CC and is capable of binding to the p62/p56lck complex to modulate lck
 CC function in a manner similar to p62. The genes transcribed in response to
 CC p160 activation likely include those of which are involved in T or B cell
 CC development/differentiation. T or B cell activation or production of T or
 CC B cell specific factors e.g. lymphokines or antibodies. This p160
 CC polypeptide is also a substrate for serine/threonine kinase activity.
 CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell
 CC cycle regulatory proteins stimulating expression of cell cycle dependent
 CC kinase inhibitors and arresting cell cycle progression at specific
 CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
 CC response it may be used to treat disorders where this is beneficial, e.g
 CC infections by pathogenic microorganisms. p160 can be used to expand T
 CC cell populations for treating infectious diseases or cancer and p160

CC inhibitors could reduce B or T cell responses and may be used to treat
CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,
CC multiple sclerosis allergic reactions, Crohn's diseases etc.
50 Sequence 3901 BP; 781 A; 1183 C; 1128 G; 809 T;

	Query Match	2.1%	Score 42.4	DB 1	Length 3901	
	Best Local Similarity	47.7%	Pred. No. 0.22%			
Matches	124	Conservative	0	Mismatches 136	Indels 0	Gaps 0
QY	1375	gaaagcctttagaaaacacgcgcgaagaaggcccggaagtgttcgtatlaaaagaactgtag	1434			
Db	3088	GAAAGGAGGAGAAGGAAGAGGAGGAGGAGGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAA	3147			
QY	1435	ataagaagaagsgyfttgaaaaagqgtggaaggttaagatactgayaagacttttt	1494			
Db	3148	GAAAGAGGGAGGAAGAGGAGGAGGAGGAGGAGCTTTGAGGAAGAGGAGAGAGAGAAAGAC	3207			
QY	1495	cagaggttgtctttaagsgtctaagggattgcaacaagatgaagagaatttctgtagggag	1554			
Db	3208	GAAATTTTTTGAAGGAGGAAGAGGAGGAGGAAGAAAGATTTCAGCAAGAAATTTAGAGGAAGAA	3267			
QY	1555	ttagagtaaagaagaagaagaatctcgaatgaactccaatggaagcaccgtgaagttgaa	1614			
Db	3268	GAAAGGTGAGTTTGAGGAAGAAAGAAAGAAAGAGAGAGTAGAGAGAGCAAGAAAGACTGGA	3327			
QY	1615	aagctttttgggagcgctt	1634			
Db	3328	GAGGTGGAAGACTCTTGAGATT	3347			

RESULT	6
ID	T89346
AC	T89346 standard; cDNA; 3211 BP.
DT	T89346:
DE	11-MAR-1998 (first entry)
KW	Human p160 cDNA 160.2.
CC	p160; p62; cytoplasmic; T cell; B cell; development; activation;
CC	modulation; cellular response; cell proliferation; autoimmune disease;
CC	p56-lck; ss.
OS	Homo sapiens.
FS	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	439..3156
FT	/*tag=a
FT	/product=p160
FT	/note="160.2"
PN	MO9722255-A1.
PD	26-JUN-1997.
PF	11-DEC-1996; U19544.
PR	10-DEC-1995; US-574959.
PA	(DAND) DANA FARBER CANCER INST INC.
PT	Joung U, Sahn U, Strominger JL, Vadlamudi RK:
DR	P-PSDB; W31186.
PT	WPI; 97-341351/3..
PT	cDNA encoding p62 and p160 and corresponding proteins - used in the
PT	treatment of autoimmune disease and for T and B cell proliferation,
PT	e.g. for treatment of tumours
PT	Claim 82, Fig 10; 175pp: English.
CC	This cDNA sequence encodes a novel p160 (160.2) which is capable
CC	of activating transcription of a variety of genes upon activation of p62
CC	and is capable of binding to the p62/p56lpx complex to modulate lck
CC	function in a manner similar to p62. The genes transcribed in response to
CC	p160 activation likely include those of which are involved in T or B cell
CC	development/differentiation, T or B cell activation or production of T or
CC	B cell specific factors e.g. lymphokines or antibodies. This p160
CC	polypeptide is also a substrate for serine/threonine kinase activity.
CC	p160 polypeptides can modulate degradation of cellular proteins e.g. cell
CC	cycle regulatory proteins stimulating expression of cell cycle dependent
CC	kinase inhibitors and arresting cell cycle progression at specific
CC	boundaries to thereby modulate cell proliferation. As p160 boosts B cell
CC	response it may be used to treat disorders where this is beneficial, e.g
CC	infections by exogenous microorganisms. p160 can be used to expand T
CC	cell populations for treating infectious diseases or cancer and p160

ID	T89346	standard; cDNA: 3211 BP.
AC	T89346	
DE	11-MAR-1998	(first entry)
DE	Human p160	cDNA 160.2
KW	p160; p62; cytoplasmic; T cell; B cell; development; activation;	
KW	modulation; cellular response; cell proliferation; autoimmune disease;	
KW	p56-lck; ss.	
OS	Homo sapiens.	
FH	key	location/Qualifiers
FT	CD5	439..3156
FT		/*tag= a
FT		/product= p160
FT		/note= "160.2"
PN	MO9722255-A1.	
PD	26-JUN-1997.	
FD	11-DEC-1996:	U19544.
FR	19-DEC-1995:	US-574959.
PA	(DAND) DNA FAREBR CANCER INST INC.	
PA	Joung I, Shin J, Strominger JL, Vadamudi RK:	
DR	WPI: 97-34151/31.	
DR	P-PSDB: W31186.	
PT	cDNA encoding p62 and p160 and corresponding proteins - used in the	
PT	treatment of autoimmune disease and for T and B cell proliferation,	
PT	e.g. for treatment of tumours	
PS	Claim 82: Fig 10; 175pp: English.	
CC	This cDNA sequence encodes a novel p160 (160.2) which is capable	
CC	of activating transcription of a variety of genes upon activation of p62	
CC	and is capable of binding to the p62/p56lck complex to modulate Lck	
CC	function in a manner similar to p62. The genes transcribed in response to	
CC	p160 activation likely include those of which are involved in T or B cell	
CC	development/differentiation, T or B cell activation or production of T or	
CC	B cell specific factors e.g. lymphokines or antibodies. This p160	
CC	polypeptide is also a substrate for serine/threonine kinase activity.	
CC	p160 polypeptides can modulate degradation of cellular proteins e.g. cell	
CC	cycle regulatory proteins stimulating expression of cell cycle dependent	
CC	kinase inhibitors and arresting cell cycle progression at specific	
CC	boundaries to thereby modulate cell proliferation. As p160 boosts B cell	
CC	response it may be used to treat disorders where this is beneficial, e.g.	
CC	infections by pathogenic microorganisms. p160 can be used to expand T	
CC	cell populations for treating infectious diseases or cancer and p160	

OY	1375	gaaaagcttggaaaacacggcggaagggcgagaagcttcttgataaagaagcttgaag	143
Db	2338	GAAGAGCGAGGAGGAAGAAGGAGAGGACGACACACACACGAAAGAGGAGAA	245
OY	1435	atagaagaagaggttgaaaaagagtgtagaaggttagagatactgaatgaattgttt	149
Db	2458	GAGCAAGCGAAGAAAGAGGAAGAGGAGAACTTTGAGGACACGGAAGAGATGAAGG	2517
OY	1495	cagaagcttgcttgaaggtttaagagattgcaacaagatgaagagaataatttgaaggag	1554
Db	2518	GAAATATTTTGAAGGAGGAAGAAAGGAGGAGGAAGAACTTTGAGGAAGATTTGAGGAGAA	2577
OY	1555	tttagtaagaagaagaagaagaattctgaatgaacttcaattgaaagcgactgaattgaa	1614
Db	2578	GAAAGCTAAGTTAGAGGAGAAAGAAAGAAAGAGAGATGACGAGAGAGAGAAAGAACTGGAA	2637
OY	1615	aagcttttgggcgcgcgctt	1634
Db	2638	GAGGTGCAACACTGAGATT	2657

Accession	Result
U01415	7
U01415	
ID	U01415 standard; cDNA; 15464 bp.
AC	U01415;
DT	17-JUN-1991 (first entry)
DE	Ryanodine receptor gene.
KW	Malignant hyperthermia; hypermetabolic syndrome; inhalation;
KW	anaesthetics; probe; calcium release channel; sarcoplasmic;
OS	reticulum; ss.
OS	Homo sapiens.

FT	key	Location/Qualifiers
FT	cds	105..15319
FT		/*tag= a
FT		/product= ryanodine receptor
FT	polya_signal	15439..15446
FT		/*tag= b
FT		15447..15457
FT		/*tag= c
FT	misc_rna	/label= Tc-rich region
FT		15464
FT		/*tag= d
FT		338..773
FT		/*tag= e
FT		/label= HRR-7B
FT		/note= "0.4 kb; claim 6"
FT	misc_rna	400..2620
FT		/*tag= f
FT		/label= HRR-7A
FT		/note= "2.2 kb; claim 6"
FT	misc_rna	2615..6359
FT		/*tag= g
FT		/label= HRR-5
FT		/note= "3.8 kb; claim 6"
FT		6349..8717
FT		/*tag= h
FT	misc_rna	/label= HRR-4
FT		/note= "2.4 kb; claim 6"
FT		9773..11075
FT	misc_rna	/*tag= i
FT		/label= HRR-3-1200
FT		/note= "1.2 kb; claim 6"
FT	misc_rna	8738..9778

PA (UNOR) Univ. of Waterloo, St. Catharines, Ont. L2N 9K1, Canada.
PA (TORO) Toronto Hospital, Toronto, Ontario, Canada.
PI (Wor) Worthington, Worthington, Ontario, Canada.
PI (WPI) 91-117517/16.
DR P-PSDB; R11510.
PT Purified DNA specific for human ryanodine receptor - useful for
PT diagnosis of malignant melanoma.
PS Claim 1: Fig 2: 49pp. English.
CC The sequence was obt'd. from several overlapping clones isolated
CC from a human skeletal muscle cDNA library in lambda gt10. The gene
CC (located on chromosome 19) encodes the human ryanodine receptor
CC (HR), a calcium release channel which spans the gap between the
CC transverse tubule and the sarcoplasmic reticulum (SR) in muscle.
CC The sustained muscle contraction in malignant hyperthermia (MH)
CC may be caused by the release of calcium into the muscle cell cyto-
CC plasm from the SR. This is due to a defect in the gene encoding
CC HR. Sequences from the gene can be used for diagnosis of MH
CC using RFLP analysis (see tags e-m).
SQ Sequence 15464 BP: 4616 C; 4735 G; 7827 T.

Query Match	2.1%	Score 42.2;	DB 1;	Length 15464;
Best Local Similarity	53.3%	Pred. No. 0.43;		
Matches	89;	Conservative	0;	Mismatches 78;
				Indels 0;
				Cons 0

[illegible]

Fragment Name	Begin	End	Accession
WP X20248_00	1	110000	
WP X20248_01	100001	210000	
WP X20248_02	200001	310000	
WP X20248_03	300001	410000	
WP X20248_04	400001	510000	
WP X20248_05	500001	610000	
WP X20248_06	600001	710000	
WP X20248_07	700001	810000	
WP X20248_08	800001	910000	
WP X20248_09	900001	910715	

Query Match	2.1%;	Score 41.2;	DB 1;	Length 9636;
Best Local Similarity	51.6%;	Pred. No. 0.65;		
Matches	94;	Conservative	0;	Mismatches 88;
				Indels 0;
				Gaps 0;

RESULT 12
V21209_00

ID V21209 standard; DNA; 1664976 BP.

Query Match	2.1%	Score 41;	DB 1;	Length 110000;
Best Local Similarity	52.0%;	Pred. No. 1.9;		
Matches 92;	Conservative	0;	Mismatches 85;	Indels 0.

RESULT	13
V74906	
ID	V74906 standard; DNA; 1484 BP.

DE	Staphylococcus aureus	conting SEQ ID #595.
KW	Computer readable medium; vaccine; S. aureus infection; immunodetection;	
KW	cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;	
KW	skin infection; surgical wound infection; scalded skin syndrome;	
KW	toxic shock syndrome; ds.	
OS	Staphylococcus aureus.	
FH	Key	location/Qualifiers
FT	misc_feature	61..120
FT	/*tag= a	
FT	/note= "these bases represent a line of missing text in	

the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence.

EP-786519-A2.
30-JUL-1997.
07-JAN-1997; 100117.
05-JAN-1996; US-009861.
(HUMA-) HUMAN GENOME SCI INC.
Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA, Rosen CA;
WPI: 97-374922/35.
Polynucleotide(s) and proteins derived from *Staphylococcus aureus* stored on computer readable medium and used in the production of anti-*S. aureus* vaccines.
Claim 1: Page 1527-1528; 3271pp; English.
This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the *S. aureus* DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against *S. aureus* infection. The polypeptides can also be used in a kit for the immunodetection of *S. aureus* in a sample. *S. aureus* is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences and their fragments are useful as primers or probes for isolating homologues of any of the *S. aureus* DNA sequences contained on the computer readable medium.
Sequence 1484 BP; 575 A; 190 C; 188 G; 469 T;

Query Match 2.1%; Score 40.8; DB 1; Length 1484;
Best Local Similarity 44.5%; Pred. No. 0.39; Indels 0; Gaps 0;
Matches 162; Conservative 0; Mismatches 202;

848 ggaattccgggtccggatcgtaatgcagtggttcaaaatttaacatgaagactta 907
580 GGAATTCCGGGTCCGGATCGTAATGCAGTGTGTTCAATATTCGACTTTTATTAAGCG 639
908 ggggaagtggtatatacaagtggttaacttaacttaacttgatgacttgatgacac 967
640 CTAAGAAAATCAAAAATTTATCAATGCTATTTCAACCTTAATTAATTCATGA 699
968 ttcatgagttcatatggaataatgataaactgttggaaacttaacatggcgataaaa 1027
700 TTATTAAGTCAAAAATACGGATGCTATGCAAGCAAAAATAAAGACACAGCAAA 759
1028 ctttgatggtgttcttctactcgatcgtcgtgcaaacattgttcaagatccagac 1087
760 ATGAGTTATTTTATTTTAAATGACTCAAAAAAGAAATTAATGAACATTAATAAATC 819
1088 ttccctggaacattataatcatgacaatgatttcttctactcaagatgactgtttaa 1147
820 TTCTTCATGAACACTATATAGTACATAGATTAATAGCCGAAATTAAGTATGTT 879
1148 tattatctcccaatcgaaaacaaacccgataattacatattcgtatatacgaagtc 1207
880 TATTGATTATATATATAGACGTGATTAATATCAATGTTTATTAATGAACGAAATTT 939
1208 gaag 1211
940 GAAA 943

RESULT 14
X13039
ID X13039 standard; DNA; 7215 BP.
AC X13039;

19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:102.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
OS vaccine; attenua ion; computer readable medium; ds.
PN Enterococcus faecalis.
MO9850555-A2.
12-NOV-1998.
04-MAY-1998; 008385.
14-NOV-1997; US-066009.
06-MAY-1997; US-044031.
16-MAY-1997; US-046655.
(HUMA-) HUMAN GENOME SCI INC.
Barash SC, Dillon PJ, Kunsch CA;
WPI: 99-045171/04.

New isolated Enterococcus faecalis polynucleotides and polypeptides used to develop products for the detection of Enterococcus and for use in vaccines for prevention or attenuation of Enterococcus infection.
Claim 1: Page 666-669; 2084pp; English.
A computer readable medium has been developed which has recorded on it 982 nucleotide sequences isolated from the Enterococcus faecalis genome. X12938 to X1319 represent these nucleotide sequences which are primary nucleotide sequences, also known as contigs. The computer-based system can identify fragments of the Enterococcus faecalis genome with commercial importance. The products can be used to detect the presence of Enterococcus faecalis in samples. They can also be used for diagnosing Enterococcal infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of Enterococcus faecalis, or another related organism, in vivo or in vitro. In particular the polypeptides encoded by the Enterococcus faecalis nucleotide sequences can be used in vaccines to prevent or attenuate an Enterococcal infection.
Sequence 7215 BP; 2685 A; 1122 C; 1520 G; 1856 T;

Query Match 2.0%; Score 39.6; DB 1; Length 7215;
Best Local Similarity 48.3%; Pred. No. 1.5; Indels 119; Gaps 0;
Matches 111; Conservative 0; Mismatches 119;

1385 agaaacagcggaagagcggaagtggttgataaagaagctgagatagaagaag 1444
5707 AAGGAGAGGCTCAACACGATGGGAAAGTAAGTAAAGTAAATAGAAAAAGAACGATC 5766
1445 agttgaaagagtggtggaaggttagagatctgaagatgacttggcagaggttc 1504
5767 GAGTGGAGAAAGACCTTAATTAAGCCAAATGACGTCGATACATTAATGAACAGTCGATG 5826
1505 ttgaaggttaagagatgcaacaagatgaagaatatttgaagggaggtgagtaag 1564
5827 CTAAGAAAAGAAATGAATGAACAAATAGAACGCGGAATTTGCTCATTAATTAACAG 5886
1565 aagagaaggaatctcgaatgaactcaaatggaagcgagctgaagttgaa 1614
5887 AAATTAATTTA-CTTTAGTAATGACTGCTGTTGAGTGAAGAAAAAGAA 5936

RESULT 15
X20389/c
ID X20389 standard; DNA; 591 BP.
AC X20389;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #142.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN MO9858943-A1.
30-DEC-1998.
18-JUN-1998; 012764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.

Thu Oct 14 07:57:24 1999

us-09-075-375-1.rst

Page 1

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: October 13, 1999, 17:41:52 ; Search time 1349.22 Seconds
(without alignments)
2896.184 Million cell updates/sec

Title: US-09-075-375-1
Perfect score: 1981
Sequence: 1 tttgggttcgaatttacc.....aattttgctaaaaaaa 1981
Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues
Database: EST:*

1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	181.2	9.1	438	24	N37612	N37612 18839 Lambd
2	145.2	7.3	602	50	A1668224	A1668224 605018D09
3	89.6	4.5	584	47	A1483033	A1483033 EST242356
4	48.6	2.5	621	35	AA550452	AA550452 1605m3 gm
5	48.4	2.4	650	41	A1055436	A1055436 coau0003P
6	42.6	2.2	802	47	A10551726	A10551726 coau0003P
7	42.6	2.2	610	51	A1723679	A1723679 BSHML3524
8	42.4	2.1	465	48	A1616040	A1616040 m66b08.Y
9	42.4	2.1	524	26	W33751	W33751 m655601.r1
10	42.4	2.1	557	38	AA771058	AA771058 vti6a09.r
11	41.1	2.1	804	45	A1324482	A1324482 m66c05.Y
12	40.8	2.1	466	39	C84248	C84248 C84248 D1C
13	40.6	2.0	766	35	AA556538	AA556538 393 L0D10
14	40.6	2.0	600	45	A1392172	A1392172 NCSM3A6T7
15	40.4	2.0	267	20	T41438	T41438 EST007 SOYB
16	40.4	2.0	571	27	AA050356	AA050356 m09d07.r
17	40.4	2.0	549	28	AA117844	AA117844 m66c05.r
18	40.4	2.0	493	30	AA239907	AA239907 m66c05.r
19	40.4	2.0	399	34	AA458418	AA458418 v66d010.r
20	40.4	2.0	419	38	AA793824	AA793824 v73c012.r
21	40.4	2.0	476	42	A1122522	A1122522 uc64q06.I
22	40.4	2.0	580	42	A1122522	A1122522 uc62d11.I
23	40.4	2.0	899	47	A0051546	A0051546 A0051546
24	40.4	2.0	274	48	A1611564	A1611564 t66b09.x
25	40.4	2.0	780	50	A0671145	A0671145 t66b09.x
26	40.4	2.0	720	50	A067145	A067145 A067145
27	40.4	2.0	483	29	AA067422	AA067422 A067422
28	40.2	2.0	419	44	A037164	A037164 A037164
29	40.2	2.0	448	26	Z74652	Z74652 MMTCC2 Mous
30	40.0	2.0	554	47	A1487591	A1487591 EST245913
31	40.0	2.0	486	47	A1488896	A1488896 EST247235
32	40.0	2.0	713	49	A0056913	A0056913 A0056913
33	39.8	2.0	360	28	C11781	C11781 C11781 yu11
34	39.8	2.0	425	28	C14744	C14744 C14744 Clon
35	39.8	2.0	324	30	AA262803	AA262803 z619a09.S
36	39.6	2.0	525	51	A1721330	A1721330 fd18e01.X
37	39.6	2.0	300	35	C53927	C53927 C53927 yu11
38	39.4	2.0	511	44	A1257228	A1257228 LP05534.5
39	39.4	2.0	748	35	AA550519	AA550519 1677m3 gm
40	39.2	2.0	300	36	C58139	C58139 C58139 yu11
41	39.2	2.0	650	39	AA870454	AA870454 vq21q12.r
42	39.2	2.0	527	43	A0034306	A0034306 A0034306
43	38.9	2.0	386	20	D33017	D33017 CELK025A5R
44	38.8	2.0	456	20	D33656	D33656 CELK034B6R
45	38.8	2.0				

ALIGNMENTS

RESULT 1
LOCUS N37612 438 bp mRNA
DEFINITION 18839 LambdA-PR12 Ataridopsis thaliana cDNA clone 207C2317, mRNA
ACCESSION N37612
NID G1158754
VERSION N37612.1 GI:1158754

05-JAN-1998

us-09-075-375-1.rst

KEYWORDS	EST.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustroids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 438)
AUTHORS	Newman,T., deBruin,F.J., Green,P., Keestra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Reitzel,E. and Somerville,C.
TITLE	Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL	Plant Physiol. 106, 1241-1255 (1994)
MEDLINE	95148729
COMMENT	On May 8, 1995 this sequence version was replaced by the following:

//strain="Arabidopsis thaliana"
 //db_xref="taxon:3702"
 //clone="207C2317"
 /clone_lib="Lambda-PR2"
 /note="Vector: lambda-PR2-Lox. Site-1: SalI. Site-2: NotI.
 Lambda PR2 is a cDNA library derived from equal
 quantities of 4 pools of mRNA. The mRNA sources were 1) 7
 day germinated etiolated seedlings; 2) tissue culture
 grown roots; 3) staged plants half dark with 24 hour light
 cycle, half on 16 hr light, 8 hour dark + rosettes; 4)
 same plants as 3 but aerial tissue (stems, flowers and
 siliques. The vector is BRL's lambda SalI-Lox. The cDNA
 inserts were directionally cloned with SalI-Mot arms using
 oligo dT primed cDNA."

Query Match	9.18	Score 181.2	DB 24	Length 438
Best Local Similarity	72.58	Pred. No. 1,1e-32		
Matches 245	Conservative	0	Matches 92	Indels 1
Qy	652	aggattgagcttcgcaaatgatatagcaaacccattcttgcyggcaaacgcttgcctgcta	711	
Db	14	AGGATGGAAGACTCGCAAAAGTCATGCGCAACCTCGCTTGAGCCAAATGCGCTGCCTT	711	
Qy	712	cagacttcgcaaatgcgttcctgacgagaccgaatgacataaatgtyggactgcttc	73	
Db	74	CAGACCTTCATTAACCGTCCAGATGAACCCGATGCCAGATTAAATGTGGGATCTGTT		
Qy	772	gaaaacagtygtgycgacccaatcaacgagtytgtygcttcccgaaagaatlytgcgcc	831	
Db	134	GAGAACGATGTTGTTGATGAGTTCACACGAGTGTGCTGTGTGGAAGAAAAGTGTGCTT	193	
Qy	832	cgyaaacgagatlybggygaaatcccggtcccgatcgtaatgagtygthtaaatltt	891	
Db	194	AGAAATCTCATCTCGAGAAATTTCTCTCCCAACACCTTCTGTCTGTGAAGAACTTC	253	
Qy	892	aactcgaagaactctgtyggaagtygatatgaacaagtygctttaaactccatcatltat	951	
Db	254	AACATCTCGACTTTTAACGGAGAGTGTACTTTCAAGTGTGCTGTAATCAACCTTGAT	313	
Qy	952	gcatttgaatttgcaactcaatgaattcatatagaaa	988	

LOCUS	A1668224	602 bp	mRNA	EST	14-MAY-1999
DEFINITION	605018DD09.x1.605--Endosperm cDNA library from Schmidt lab Zea mays.				
ACCESSION	AF1668224				
NID	54827532				
VERSION	A1668224.1				
KEYWORDS	EST.				
SOURCE	Zea mays.				
ORGANISM	Zea mays.				

Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 605018 row: D column: 09.
 Location/Qualifiers
 1. 602
 source

```
Query Match      7.38; Score 145.2; DB 50; Length 602;  
Best Local Similarity 63.5%; Pred. No. 3.2e+24;  
Matches 244; Conservative 0; Mismatches 128; Indels 12; Gaps 1  
1272 attcgatcattccaccccc
```

Db 600 AACAAATAAACCGGAGCTGTGAAGAGCGCAAAAGCGTAGTGCGGACTTTCGACGTT 541
 QY 1332 cataaacacgcgcgaatagtgttggtgscgtgaagctcaattgtgtgaaaggcttgagaanaac 1391
 Db 540 CATCATGACCGGCACAAACACCTGTGCTTGAGCCTCCTCTTGTGGAGAGATGAGAAAC 481
 QY 1392 aacgggaagaaggcggaagaagttgttgabaaagaagctgtagagataagaagaaggttga 1451
 Db 480 TGTGTGGAAGGAGAGAACCCATCTGTCAGGGAGGTGAAGAGATCGAGAGACGATTGA 421
 QY 1452 aaagaaggttggaagggttagagatactagatgaacttggttcagaaggttgcgtgaag 1511
 Db 420 GGAGCTGGAGAGAGAG-----GAGCGTCACTGTTTCAGAAAGCGGCAGAGAG 373
 QY 1512 gtttaagaagcttcaacaagaatgaagaagaatttgttgaggagtttgaaagaagaa 1571
 Db 372 TCTCATGAGAGCTCAAAACAGAGATTTTCATGAACCTCTTCGAGGGCGCTGAGCAAGAGAGGAT 313
 QY 1572 ggaattctgaatgaacttcaaatggaagcgactgaagttgaaaggtttttgggcgcgc 1631

QY	1342	gacataagttctgyccttgacgctccatctgtgtgaaagcgttgaagaacagcggaagag	1401
Db	158	GAAATAAGTGCATGATGAAGTACGAAAGAAATGTGATGAAGAAGATGAGAA	217
QY	1402	gagcgagaagttcttgataaagaagcgtgtagagatagagaagaagctgtgaagaagagctg	1461
Db	218	GATGAATAATGATTAAGATGATGTGGAGGAAGGAGAAAGATGATGTACAAGAACGAA	277
QY	1462	gagaaagttcagagtaactcagagatgactctgttcacagaagctgtcgtgaagggtttaaggaa	1521
Db	278	GACCAAGACAAATGATGATGAGAAAGATGTAAGATTTTATGTTGAAAGAAATCTATCGAA	337


```

FEATURES
    source
        i. 466
        /organism="Dictyostelium discoideum"
        /strain="AX4"
        /db_xref="taxon:44689"
        /clone="SSB385"
        /clone.lib="Dictyostelium discoideum SS (H.Urushinara)"
        /dev_stage="slug"
BASE COUNT      270 a      35 c      56 g      152 t      3 others
ORIGIN

```

Query Match

2.1%; Score 40.8; DB 39; Length 466;
 Best Local Similarity 46.2%; Pred. No. 9.3;
 Matches 132; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 1526 aacaagatgaaggaatttctgtgagggagttgagtaagaagaaggaaggaattcgaatg 1585
 |||||
 Db 124 AACAGAGAGTTGTAAGAGAAATTAAGAAATATGTTAATGGGTAGTAATTCGATTAAT 183
 QY 1586 aactcaaatggaagcagcagtaagttgaaagcgttttgagcgcggtaccgattagga 1645
 |||||
 Db 184 CACAGAGAAATGTAATTTGATGAGAGTGTATATGTAATTCGTAATCAAAACCAATTTTA 243
 QY 1646 aacttgataaattcgaatgattgattcagacaatatatagtcataatgattatgtag 1705
 |||||
 Db 244 ATATAGACCAATTTTATGATTAATAATCGAACGCTGTAATTAATTAATTAATTAATA 303
 QY 1706 atactgagagaaccccaaaacatttgatatacgtgataaagctgttgattgattgtt 1765
 |||||
 Db 304 CAATTAATATATATATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 363
 QY 1766 atgagcttaaaattgagaatagcttttaattcttcaaaaaa 1811
 |||||
 Db 364 TATTAATATAGATTTTAAAGAAAGTATTAATTAATTAATTAATTAATTAATTAATTA 409

RESULT 13

AA556538 766 bp mRNA EST 28-AUG-1998
 LOCUS 393 Lobolly pine C Pinus taeda cDNA clone 3C2D, mRNA sequence.
 DEFINITION
 AA556538 9335553
 NID
 VERSION AA556538.1 GI:3365553
 KEYWORDS
 EST.
 SOURCE lobolly pine.
 ORGANISM Pinus taeda
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Coniferopsida; Coniferales;
 Pinaceae; Pinus

REFERENCE
 1 (bases 1 to 766)
 Allona, I., Quinh, M., Shoop, E., Swope, K., St. Cyr, S., Carlis, J.,
 Riedel, J., Retzel, E., Campbell, M. M., Sedoreff, R. and Whetten, R. W.
 Analysis of xylem formation in pine by cDNA sequencing
 Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9693-9698 (1998)
 98356220
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1339611.

CONTACT: Ross Whetten
 Forest Biotechnology Group
 North Carolina State University
 Dept. of Forestry, NC State University, 6113 Jordan Hall,
 Raleigh, NC 27695-8008
 Tel: 919-515-7800
 Fax: 919-515-7801
 Email: rosswhetten@unity.ncsu.edu
 Seg primer: T3.

FEATURES
 source
 1. 766
 /organism="Pinus taeda"
 /strain="Coastal plain lobolly pine from North Carolina"
 /db_xref="taxon:3352"
 /clone="3C2D"
 /clone_lib="lobolly pine C"
 /tissue_type="Xylem"
 /lab_host="SOLR"
 /note="Vector: lambda-ZAP. Site.1: EcoRI; Site.2: XhoI;
 The library was made from immature xylem from the
 underside of inclined stems of differentiating wood. A
 mixture of four genotypes were used. Oligo-dt prime cDNA
 was directionally cloned into the EcoRI-XhoI lambda-ZAP
 vector arms"

BASE COUNT 230 a 132 c 182 g 199 t 23 others
 ORIGIN

Query Match 2.0%; Score 40.6; DB 35; Length 766;
 Best Local Similarity 48.7%; Pred. No. 11;
 Matches 134; Conservative 0; Mismatches 138; Indels 3; Gaps 1;

QY 1382 ttgagaacacagcggaagagcgagagctgttgataaagaagctgaagatgaag 1441
 |||||
 Db 202 TAGAGTATGCAACAAAGCTGTGATTAATGCGGAAGCTGATTTGGCAATTAAGTCCAAAG 261
 QY 1442 aagagctgaagaagagctgagagattgagatactgagattgcttcaagaggt 1501
 |||||
 Db 262 ATGGGATTTGTTTGGGGGTGAGAGAGCTCATTCCTCTTAAGAAG---CTTCTTAANAGAT 318
 QY 1502 tgcctgaagggattgaagagctgcaacaagaatgaagaattttgagagagtgagta 1561
 |||||
 Db 319 CCAATCGAAGATTCATGCTGTTCACATCATCTGTCGATGGGTGCTGCTGATTCGACAG 378
 QY 1562 aagaagaaggaagattctgaatgaacttaaatggaagcagctgaagttgaagctt 1621
 |||||
 Db 379 CAGATGATGACCAATTTGTGCTGCTTAATCAACAGCAAGCAAGTTATGANAACACTT 438
 QY 1622 ttggcgcgcttaccgattaggaactagataa 1656
 |||||
 Db 439 ATGGTGAAGCCGATCCCTATTAGGAACCTTGTGAA 473

RESULT 14

AI392172 600 bp mRNA EST 03-FEB-1999
 LOCUS AI392172/C
 DEFINITION
 NC036677 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 3'
 similar to uridine kinase, S. cerevisiae, mRNA sequence.
 ACCESSION
 AI392172
 NID
 VERSION AI392172.1 GI:4219979
 KEYWORDS
 EST.
 SOURCE Neurospora crassa.
 ORGANISM Neurospora crassa
 Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
 Sordariales; Sordariaceae; Neurospora.

REFERENCE
 1 (bases 1 to 600)
 Nelson, M. A., Kang, S., Braun, E. L., Crawford, M. E., Dolan, P. L.,
 Leonard, P. M., Mitchell, J., Armijo, A. M., Bean, L., Bueyes, E.,
 Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K.,
 Miller, R., Ortega, J., Pavlov, J., Perea, J., Todisco, S.,
 Trujillo, R., Valentine, J., Wells, A., Werner, Washburne, M., Yazzie, S.
 and Natvig, D. O.
 Expressed sequences from conidial, mycelial, and sexual stages of
 Neurospora crassa
 Fungal Genet. Biol. 21, 348-363 (1997)
 97435549
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1796934.

CONTACT: Natvig, D. O./Nelson, M. A.
 Department of Biology
 University of New Mexico
 Castetter Hall, Albuquerque, NM 87131, USA
 Tel: 505 277 3411
 Fax: 505 277 0304
 Email: ngp@biology.unm.edu.
 FEATURES
 source
 1. 600
 /organism="Neurospora crassa"
 /strain="74-OR23-IV A (FGSC 2489)"
 /db_xref="taxon:5141"
 /clone="SM3A6"
 /clone_lib="Subtracted Mycelial"
 /sex="Mating type A"
 /tissue_type="Mycelium"
 /dev_stage="Mycelium"
 /lab_host="E. coli"
 /note="Vector: pBluescript SK (-); Site.1: EcoRI; Site.2:
 XhoI; 2% sucrose for 24 hours. cDNA directionally cloned

BASE COUNT 230 a 132 c 182 g 199 t 23 others
 ORIGIN

into pBluescript SK(-) using the Uni-ZAP XR vector system (Stratagene, La Jolla, CA). Previously identified highly expressed clones were subtracted from this library."

BASE COUNT 125 a 215 c 102 g 158 t

Query Match 2.0%; Score 40.6; DB 45; Length 600;
Best Local Similarity 64.2%; Pred. No. 11;
Matches 61: Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1397 aagagggcagagagtggtgtgataaaagagctgtcagagatgaagaagaggttcaaaag 1466
|||||
Db 215 AAGAGGTATTTCAGGTGTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 156

QY 1457 aggtgagagaggttagagactgagactgtg 1491
|||||
Db 155 AGATGCTGATGCTCATTCATGATGAGAGTGTTC 121

RESULT 15

T41438 267 bp mRNA EST 20-JUL-1995
LOCUS EST007 Soybean, M Bhattacharya Glycine max CDNA clone SPM1 5',
DEFINITION mRNA sequence.
T41438
ACCESSION
NID 9957381
VERSION T41438.1 GI:957381
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
Glycine.

REFERENCE 1 (bases 1 to 267)
AUTHORS Shi, J., Dixon, R.A., Gonzales, R.A., Kjellbom, P. and
Bhattacharya, M.K.
TITLE Identification of CDNA clones encoding valosin-containing protein
and other plant plasma membrane associated proteins by general
immunoselecting strategy
Proc. Natl. Acad. Sci. U.S.A. 92, 4457-4461 (1995)
95273382

COMMENT

Other ESTs: EST060
Contact: Madan Bhattacharya
Plant Biology Division
The Samuel Roberts Noble Foundation, Inc.
2510 Sam Noble Parkway, Ardmore, OK, 73401
Tel: 4052235810
Fax: 4052217380
Email: madan@ardvark.ucs.uoknor.edu
Seq primer: SK.

FEATURES

source
1..267
Location/Qualifiers
/organism="Glycine max"
/strain="cv Williams 82"
/db_xref="taxon:3847"
/map="21"
/clone="SPM1"
/note="Vector: Uni-ZAP. Site1: EcoRI; Site2: XhoI; mRNA
was purified from soybean etiolated hypocotyls. CDNA was
constructed using an XhoI/oligo (dT)-linker/primer and
cloned unidirectionally into the EcoRI and XhoI sites of
the lambda phage Uni-ZAP (Stratagene). The library was
screened using an anti-plasma membrane serum preabsorbed
with an E. coli/phage lysate."
BASE COUNT 123 a 28 c 90 g 26 t
ORIGIN

Query Match 2.0%; Score 40.4; DB 20; Length 267;
Best Local Similarity 50.5%; Pred. No. 11;

Matches 98; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1385 agaaacagcgcagagggcgagagatgtgtatataaagagcttgaagataagaag 1444
|||||
Db 46 AGAACCAGCAGCAG 105

QY 1445 aggttcaaaagaggtgagagaggttagagactgagactgttcaagaggttc 1504
|||||
Db 106 AACTCATAGAAAG 165

QY 1505 ttgaaggttttagagaggttgcacaaagatgaagagatlttgtagggagttgaag 1564
|||||
Db 166 CAG 225

QY 1565 aagagagagagat 1578
|||||
Db 226 AGAAGAGAGAGAGAT 239

Search completed: October 13, 1999, 18:23:30
Job time: 2498 sec

Thu Oct 14 07:57:24 1999

us-09-075-375-2.rag

Page 1

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 1999, 03:11:17 ; Search time 34.13 Seconds
(without alignments)
979.923 Million cell updates/sec

Title: US-09-075-375-2
Perfect score: 7495
Sequence: 1 MALSLHTVFLCKEALNLVLA.....MEASEVEKLFGRALPIRKVR 1412

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2554	34.1	478	1	Tobacco violaxanthin
2	2498	33.3	473	1	Romaine lettuce vi
3	2417	32.2	462	1	Arabidopsis violax
4	147.5	2.0	1786	1	P. falciparum live
5	147	2.0	2482	1	Human mitotin amln
6	147	2.0	2482	1	Human mitotin amln
7	135.5	1.8	1312	1	Human homologue of
8	135.5	1.8	1312	1	Human RAD50. Human
9	135	1.8	968	1	Human p85alpha-green flo
10	134	1.8	1639	1	P. falciparum synt
11	132	1.8	776	1	Lethal factor of B
12	129.5	1.7	2954	1	Amino acid sequenc
13	129	1.7	1654	1	Sequence of the p1
14	129	1.7	993	1	Pre-nisin modifica
15	124	1.7	1582	1	Rat sulphonylurea
16	124	1.7	1498	1	Rat sulphonylurea
17	121.5	1.6	1132	1	Chicken leucocytos
18	121	1.6	1411	1	Nucleolar/endosoma
19	119.5	1.6	1604	1	TNF-R-EBA 175 fus1
20	119.5	1.6	1099	1	Mutant C-beta prot
21	119	1.6	1135	1	Human p160 polypep
22	119	1.6	905	1	Human p160 polypep
23	118.5	1.6	2273	1	Erysiphe graminis
24	118.5	1.6	1130	1	Merostin major subu
25	115.5	1.5	3110	1	Merostin major subu
26	115.5	1.5	1435	1	P. falciparum SABP
27	115	1.5	1435	1	Siliate acid bindin
28	114.5	1.5	1230	1	Potato tuber solub
29	114.5	1.5	1743	1	* H. pylori GHPO 175
30	114.5	1.5	3248	1	Kinetochore protei
31	112	1.5	649	1	Expressed antigen
32	111.5	1.5	1130	1	Merostin M polypept
33	111.5	1.5	2329	1	Partial BRCA2 canc
34	111.5	1.5	3418	1	Human breast and o
35	111.5	1.5	3418	1	Human BRCA2 (om14)
36	111.5	1.5	3418	1	Human BRCA2 (om11)
37	110.5	1.5	1786	1	CD4-EBA175 fusion
38	110.5	1.5	1164	1	Group B Streptococ
39	110.5	1.5	1164	1	Group B Streptococ
40	110.5	1.5	1128	1	Mutant C-beta prot
41	110.5	1.5	1093	1	Mutant C-beta prot
42	110.5	1.5	1164	1	Mutant C-beta prot
43	110	1.5	751	1	H. pylori GHPO 325

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB ID	Description
1	W09875	110	1.5	1392	1	Tobacco violaxanthin de-epoxidase
2	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
3	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
4	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
5	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
6	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
7	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
8	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
9	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
10	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
11	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
12	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
13	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
14	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
15	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
16	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
17	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
18	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
19	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
20	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
21	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
22	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
23	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
24	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
25	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
26	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
27	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
28	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
29	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
30	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
31	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
32	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
33	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
34	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
35	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
36	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
37	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
38	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
39	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
40	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
41	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
42	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase
43	W09875	109.5	1.5	933	1	Arabidopsis violaxanthin de-epoxidase

CC (T66242). VDE nucleic acids (see also T66241, T66243), in sense
 CC or antisense orientation, can be used in genetic constructs to
 CC modify VDE levels in plants. Increased levels result in the plant
 CC being tolerant of increased light and therefore more productive
 CC and/or more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 S0 Sequence 478 AA;

Query Match 34.1%; Score 2554; DB 1; Length 478;
 Best Local Similarity 99.8%; Pred. No. 1.6e-193;

Matches 476; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 474 MALAPSNFLANHETIKYVGSKLPGHRRFSMGWEDYFGSIYAKICSSRIPIRYRKPSP 533
 DB 1 MALAPSNFLANHETIKYVGSKLPGHRRFSMGWEDYFGSIYAKICSSRIPIRYRKPSP 60
 QY 534 RICCGLDSRGDILFSGHKNLSPASINQNPKNSSCKPKDYALMWEKMGQFAKTAI 593
 DB 61 RICCGLDSRGVOLFSGHKNLSPASINQNPKNSSCKPKDYALMWEKMGQFAKTAI 120
 QY 594 VAFILISVASKADAVDAKLTCTCLKECLAKCISNPACANVACLOTNNRPDETEC 653
 DB 121 VAFILISVASKADAVDAKLTCTCLKECLAKCISNPACANVACLOTNNRPDETEC 180
 QY 654 QIKCGDLFENSVDENCAVSRKCYPRKSDVDGDFPVPDPSVLVOKFDMKDFSGKWEIT 713
 DB 181 QIKCGDLFENSVDENCAVSRKCYPRKSDVDGDFPVPDPSVLVOKFDMKDFSGKWEIT 240
 QY 714 RGLNTPFAFPCQJHEFHTEENKLVGNLSWIRTPDGGFFTSAYOKFVODKXPGYILYN 773
 DB 241 RGLNTPFAFPCQJHEFHTEENKLVGNLSWIRTPDGGFFTSAYOKFVODKXPGYILYN 300
 QY 774 HDNEYLLODDWYLLSSKVENSPEDYIFVYKGRNANMGVGSVLYRSVALESTIPE 833
 DB 301 HDNEYLLODDWYLLSSKVENSPEDYIFVYKGRNANMGVGSVLYRSVALESTIPE 360
 QY 834 LQTAOKVNGDENTFIKTNTCGPEPLVERLEKKEEGEERTIIKEVEIEEVEKVRDK 893
 DB 361 LQTAOKVNGDENTFIKTNTCGPEPLVERLEKKEEGEERTIIKEVEIEEVEKVRDK 420
 QY 894 EVTLFSKLFEGFKELQDENFLRELSKREPMVLLGLMEATEVEKLFGRALPIRLK 950
 DB 421 EVTLFSKLFEGFKELQDENFLRELSKREPMVLLGLMEATEVEKLFGRALPIRLK 477

RESULT 2

ID W09874 standard; Protein: 473 AA.
 AC W09874;

DE 28-JUL-1997 (first entry)
 KM Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KM photoprotection; transgenic plant; zeaxanthin; anthraxanthin;
 OS Lactuca sativa L. cv. romaine.

EH Key location/Qualifiers
 FT 1..125

FT peptide /label= Transit-peptide
 FT 126..473

FT protein /label= Mat.protein
 FT 126..138

FT peptide /note= "Clain 8"
 FT 126..197

FT domain /label= Cys-rich domain
 FT 218..231

FT peptide /label= Lipocalin signature
 FT 376..473

FT domain /label= Highly-charged domain
 FT 265..272

FT peptide /label= tryptic-peptide-11
 FT 275..289

FT FT /label= Tryptic-peptide-21
 FT peptide 341..353
 FT FT /label= Tryptic-peptide-15
 FT misc_difference 132
 FT /note= "conserved Cys residue"
 FT misc_difference 134
 FT /note= "conserved Cys residue"
 FT misc_difference 139
 FT /note= "conserved Cys residue"
 FT misc_difference 146
 FT /note= "conserved Cys residue"
 FT misc_difference 152
 FT /note= "conserved Cys residue"
 FT misc_difference 158
 FT /note= "conserved Cys residue"
 FT misc_difference 162
 FT /note= "conserved Cys residue"
 FT misc_difference 171
 FT /note= "conserved Cys residue"
 FT misc_difference 175
 FT /note= "conserved Cys residue"
 FT misc_difference 190
 FT /note= "conserved Cys residue"
 FT misc_difference 197
 FT /note= "conserved Cys residue"
 FT misc_difference 243
 FT /note= "conserved Cys residue"
 FT misc_difference 373
 FT /note= "conserved Cys residue"
 FT misc_difference 373
 FT /note= "conserved Cys residue"
 FT W0917447-A2.
 PD 15-MAY-1997.
 PE 07-NOV-1996; US-8291.
 PR 07-NOV-1995; US-006315.
 PR 06-AUG-1996; US-023502.
 PA (CALJ) CALGENE INC.
 PI Bugs RC, Rockholm DC, Yamamoto Hy;
 DR WPI: 97-281036/25.
 DR N-PSDB: T66241.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 PS Example 1; Fig 1; 4pp; English.
 CC The 55 kDa violaxanthin de-epoxidase (VDE) (W09874) of romaine
 CC lettuce catalyses the de-epoxidation of violaxanthin to zeaxanthin
 CC and anthraxanthin. This system, termed energy dependent
 CC non-radiative energy dissipation or non-photochemical fluorescence
 CC quenching, reduces the quantum efficiency of photosystem II (PSII)
 CC helping to prevent psii over-reduction and photoinhibitory damage.
 CC The amino acid sequence of the VDE was deduced from an isolated
 CC cDNA clone (T66241). VDE nucleic acids (see also T66242-43), in
 CC sense or antisense orientation, can be used in genetic constructs
 CC to modify VDE levels in plants. Increased levels result in the
 CC plant being tolerant of increased light and therefore more
 CC productive and/or more resistant to disease. Underexpression of
 CC VDE increases photosynthetic efficiency under low light. The
 CC photosensitivity of a range of crops, trees and ornamentals can be
 CC modified.
 S0 Sequence 473 AA;

Query Match 33.3%; Score 2498; DB 1; Length 473;
 Best Local Similarity 99.6%; Pred. No. 4.1e-189;
 Matches 471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MALSHYFLCKEALMLVYARSPCNERFHRSGOPPTNIMKIRSNMGYFNSFRLEFSYK 60
 DB 1 MALSHYFLCKEALMLVYARSPCNERFHRSGOPPTNIMKIRSNMGYFNSFRLEFSYK 60
 QY 61 TSFSDSHCKDKSQICSDTSPFEIQRDILRGMTLLIEKOWROFOLATLVCTEVIV 120
 DB 61 TSFSDSHCKDKSQICSDTSPFEIQRDILRGMTLLIEKOWROFOLATLVCTEVIV 120
 QY 121 PRDVAVALTTCACLLKECRTELAKCIANPSCANVACLOTNNRPDETEOICGGLFE 180

Db 121 PRDVAVALKTCACLKECHIELAKCIANSCANVACLOTNNRPDETECOIKGDLFE 180
 QY 181 NSVDQNECAVSRKRCVPRKSDVGEPPVDRAVAVONENKMDPSKWTITGLNPTDA 240
 Db 181 NSVDQNECAVSRKRCVPRKSDVGEPPVDRAVAVONENKMDPSKWTITGLNPTDA 240
 QY 241 FDCOLHEFHENDKLVGNLWRIKTDGGFTSAVOTFYQDDPLGALYNDNEFLHQ 300
 Db 241 FDCOLHEFHENDKLVGNLWRIKTDGGFTSAVOTFYQDDPLGALYNDNEFLHQ 300
 QY 301 DDWVILSSQIENKPDYIFVYRRNDAMDYGSGSVIYTRSPILPESIIPILOKAAKSVG 360
 Db 301 DDWVILSSQIENKPDYIFVYRRNDAMDYGSGSVIYTRSPILPESIIPILOKAAKSVG 360
 QY 361 RDNFNFITDSCGPEPPLVERLEKTAEBEGKLLIKAVITEEVEKEVKADTMTLF 420
 Db 361 RDNFNFITDSCGPEPPLVERLEKTAEBEGKLLIKAVITEEVEKEVKADTMTLF 420
 QY 421 QRLLEGFELOODEENFVRELKREKELINLOMEATEVEKLGALPIRKL 473
 Db 421 QRLLEGFELOODEENFVRELKREKELINLOMEATEVEKLGALPIRKL 473
 RESULT 3
 ID W09876
 AC W09876 standard: Protein; 462 AA.
 DT 28-JUL-1997 (first entry)
 DE Arabidopsis violaxanthin de-epoxidase.
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KM photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KM xanthophyll.
 OS Arabidopsis thaliana var. columbia.
 FH Key Location/Qualifiers
 FT peptide 1..113
 FT /label=Transit-peptide
 FT 114..462
 FT /label=Mat-protein
 FT 114..126
 FT /note="Claim 8"
 FT 114..185
 FT domain
 FT /label=Cys-rich_domain
 FT 364..462
 FT domain
 FT /label=Highly-charged_domain
 FT 120
 FT /note="conserved Cys residue"
 FT 122
 FT /note="conserved Cys residue"
 FT 127
 FT /note="conserved Cys residue"
 FT 134
 FT /note="conserved Cys residue"
 FT 140
 FT /note="conserved Cys residue"
 FT 146
 FT /note="conserved Cys residue"
 FT 150
 FT /note="conserved Cys residue"
 FT 159
 FT /note="conserved Cys residue"
 FT 163
 FT /note="conserved Cys residue"
 FT 178
 FT /note="conserved Cys residue"
 FT 185
 FT /note="conserved Cys residue"
 FT 231
 FT /note="conserved Cys residue"
 FT 362
 FT /note="conserved Cys residue"
 PN W0917447-A2.
 PD 15-MAY-1997.
 PF 07-NOV-1996, U18291.

PR 07-NOV-1995; US-06315.
 PR 06-AUG-1996; US-C23502.
 PA (CALJ) CALGENE INC.
 PI Bugos RC, Rockfollm DC, Yamamoto HY.
 DR WPI; 97-281036/25.
 DR N-PSDB; T66243.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 PS Disclosure; Fig 3; 41pp; English.
 CC The violaxanthin de-epoxidase (VDE) (W09876) of Arabidopsis
 CC catalyses the de-epoxidation of violaxanthin to zeaxanthin and
 CC antheraxanthin. This system, termed energy dependent non-radiative
 CC energy dissipation or non-photochemical fluorescence quenching,
 CC reduces the quantum efficiency of photosystem II (PSII), helping to
 CC prevent PSII over-reduction and photoinhibitory damage. The amino
 CC acid sequence of the VDE was deduced from an isolated cDNA clone
 CC (T66243). VDE nucleic acids (see also T66241-42), in sense or
 CC antisense orientation, can be used in genetic constructs to modify
 CC VDE levels in plants. Increased levels result in the plant being
 CC tolerant of increased light and therefore more productive and/or
 CC more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 SQ Sequence 462 AA:
 Query Match 32.2%; Score 2417; DB 1; Length 462;
 Best Local Similarity 99.8%; Pred. No. 9, 9e-183;
 Matches 461; Comparative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 951 MAVATHCFTSPCHDRIRFFSSDDIGRLGTRKRNGTFLKILPIPGADIRTTGRRSS 1010
 Db 1 MAVATHCFTSPCHDRIRFFSSDDIGRLGTRKRNGTFLKILPIPGADIRTTGRRSS 60
 QY 1011 RPLSAFRSGFSKGIPIVLPSPKNEKELTAPLLKLVGLACAFILVPSADVDALKTC 1070
 Db 61 RPLSAFRSGFSKGIPIVLPSPKNEKELTAPLLKLVGLACAFILVPSADVDALKTC 120
 QY 1071 ACLKGCRIEYAKCIANPACANVACLOTNNRPDETECOIKGDLFENSVDENECAV 1130
 Db 121 ACLKGCRIEYAKCIANPACANVACLOTNNRPDETECOIKGDLFENSVDENECAV 180
 QY 1131 SRKCCVPRKSDIGEPAPDPSPVLVONFNISDENGKWTITGLNPTDAPDQCLHEHTHG 1190
 Db 181 SRKCCVPRKSDIGEPAPDPSPVLVONFNISDENGKWTITGLNPTDAPDQCLHEHTHG 240
 QY 1191 DNKLGNISMRKTLDSGFTFRSAVQKRVODPNQGVLYNHNDNEYLYHODDWIISKIE 1250
 Db 241 DNKLGNISMRKTLDSGFTFRSAVQKRVODPNQGVLYNHNDNEYLYHODDWIISKIE 300
 QY 1251 NKPEDIIFVYRRNDAMDYGAGVYTRSSVLPNSIIPLEKAAKSIGDFSTFRTDN 1310
 Db 301 NKPEDIIFVYRRNDAMDYGAGVYTRSSVLPNSIIPLEKAAKSIGDFSTFRTDN 360
 QY 1311 TCGPEPALVERIEKTEVEGERIIIVKEVEEIEEVEKEVEKVGRTMTLFORLAEGNEIK 1370
 Db 361 TCGPEPALVERIEKTEVEGERIIIVKEVEEIEEVEKEVEKVGRTMTLFORLAEGNEIK 420
 QY 1371 QDEENFVRELKREKELINLOMEATEVEKLGALPIRKL 1412
 Db 421 QDEENFVRELKREKELINLOMEATEVEKLGALPIRKL 462
 RESULT 4
 ID W24790
 AC W24790 standard: Protein; 1786 AA.
 DT 08-OCT-1997 (first entry)
 DE P. falciparum liver stage antigen-3.
 KW Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
 KM Propylaxis; Thal strain; gene organisation; exon; intron; hydrophobic;
 KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;
 KW vaccine; immunotherapy; malaria.

OS Plasmodium falciparum.
 FH Key Location/Qualifiers
 FT region 223..278
 FT /note= "repeat region 1"
 FT region 279..818
 FT /note= "repeat region 2"
 FT 1537..1576
 FT /note= "repeat region 3"
 PN WO9641877-A2.
 PD 27-DEC-1996.
 PF 12-JUN-1996; F00894.
 PR 13-JUN-1995; FR-007007.
 PA (INSP) INST PASTEUR.
 PI Daubersies P, Druilhe P;
 DR N-PSDB: 778868.
 PT Plasmodium falciparum poly:peptide(s) and related nucleic acids -
 PT derived from the liver stage antigen-3, useful for malaria vaccine
 PT prodn. and diagnosis
 PS Claim 1; Fig 2A-I; 69pp; French.
 CC This sequence corresponds to a Plasmodium falciparum strain K1
 CC pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding
 CC gene sequence was isolated by screening a P. falciparum strain 19/96
 CC library with serum from a missionary treated by prophylaxis (for strain
 CC T6/96 see FR101286). Of 20 clones isolated, clone 7295 was used to
 CC screen a library generated from Thai strain K1. One clone contained a
 CC 6.85 kb insert including the genomic sequence 178867. The gene comprises
 CC a 1.8 kb region encoding 3 major blocks of tetrapeptide repeats
 CC (especially the amino acid sequence VEEV, VEEV, VAPV, VAPV, etc)
 CC and a 3' hydrophobic region corresponding to a glycosyl-phosphatidyl-
 CC inositol membrane anchoring sequence. The invention relates to new
 CC polypeptides of at least 10 amino acids derived from the LSA-3 protein
 CC with the exception of the peptides W24791-4. The LSA-3 peptides can be
 CC used to raise antibodies and as vaccines for immunotherapy of malaria.
 SQ Sequence 1786 AA;

Query Match 2.08; Score 147.5; DB 1; Length 1786;
 Best Local Similarity 18.18; Pred. No. 0.013;
 Matches 207; Conservative 155; Mismatches 374; Indels 409; Gaps 50;

QY 347 SIIFNLOK-AKSVGRDFN-----NFITDNSCGPEPLVE-----RLEKTAENG 390
 DB 858 TILNVEETIAESTTTSNILEIQTNTINDIT---EKLLEPHENYLSALENTQSE 914
 QY 391 EKLITKAVELLEVEKEV-----EKVDTENTLFORLLEGEKLEQD-----EENFVEL 441
 DB 915 EK--KEVIDVIEEKEEVATLIEVEQAEKSNMTIEIFENLENNVESNENVAENTL 971
 QY 442 SKEKEITLNEQATEVEKLFGRALPIRKLNALAPHSNFLANHETTYVYGSKLPGHK 501
 DB 972 EKLNETYFNVLKVEETVIEISGSELENMEMDAF-----FSEITDNVGIQENILTG-- 1025
 QY 502 RFSWGMEDYGSIVYAKICSSRIIPRYFKSPICGLDSRGHOLFESH-----KHNLEPA 557
 DB 1025 -----MFRSIETSIIVIGSE--KYVDLENVYSSIIDN--IENKKEBLKNLENIEST 1072
 QY 558 HSNINONPKG-NSGCKFPKVALMWEKGOFAKTAIVAIFLISVASKRADAVDLKTCCTC 616
 DB 1073 EGVQETVTEHVEQNYVYDVDPAMK-----DOF-----LGLINENGK----- 1110
 QY 617 LKKECLEIAKICSNPACANVACLOTNNRPDETECO---IKCDDLTENSVDENFENCA 673
 DB 1110 -LKEMFENLEDVFKS---ESDVTIVTEFKDEPVQKEVEKETYSIIEMEENIVDLLE-- 1164
 QY 674 VSRKKCVPRKSDVD--FPVPPSPVLVOKFDKDKSGKWFITRGINPTFADFDCQLHFEH 731
 DB 1164 -----EKEDLTDMIDAVESEIIS--DSKEET-----EST 1193
 QY 732 TEENKLVGNLSMIRTPDGGFTFRSAVOKFVDDPKYPGILVHNHNEYLLYDDDWYIISSK 791
 DB 1194 KDKKDVSLIVEFVQND-----MDESVEKYLE-----LKNMEBELMKDAVEINDITRSK 1242

QY 792 -VENSPEYITFYVYKGRNDAMDGYGGSVLYTFRSAVLPESI--IPELOTAAQVGRDFTNF 848
 DB 1243 LIETQI-----LNEVADLITKMEKLEKKA-----LSDSNEI 1278
 QY 849 IKTNCTGPEPPVLEVERLEKKEVEGE--RTIKVEEI-----EEVEKVRD-----K 893
 DB 1279 IDAD-----DTLEKVEEHDITTLDEVLVVKLQVEEDKTEKSDKLEDEIDILK 1329
 QY 894 EV-----TLFKLEEGFKELORDEENFLRELSKEMVDLDGLMEATEVEKLFGRAL----- 946
 DB 1330 EVKEIKLESEIIEEDYKELKITFTDLEKKEIKENKHFKEEAELEIKDLADILKEVS 1389
 QY 946 -----PIRKMAVAITHCFSPCHDRIRFFSSDGIQRLGIRKRNGIFFLKIL 994
 DB 1390 SLEVEEKKLEVEHKEVEEYHIIISGDH-----IKG----- 1422
 QY 995 PIIOADLRTTGGRSSRPLSAFRSGSKGIFDIPLPSKNEIKELTAPLLKLINGVLACA 1094
 DB 1422 -LEEDDLEEVDDLKSLIDMLKGMELGMD-----KESLEDVTKL----- 1463
 QY 1055 FLIYPSADAVDAKLTACALIKGRIELACIANPACANVACLOTGN--NRPDETECOIK 1112
 DB 1463 -----GERVESLKDVLSSALGMDEBQMKTKKAQPKLEEVILK 1501
 QY 1113 CGDLFENSVDENNECAVARRKCVPRKSDLGFFPAPDSVLVQNNNISDFNGKWTITSG 1172
 DB 1502 -----EEVKEPKRKRTKKR-----VRDID-----K 1524
 QY 1173 NPTEFADFCDLHBFHTEGDNKLGVNISMRIKTLDSGFTFRSAVOKFVDDPQNGVLYNHD 1232
 DB 1525 EPKDEIYEVEMKQEDIEED-----VEEDIED----- 1552
 QY 1233 NEVLHQQDDMYLISSKYENKPEPDYIFVYRGRNDAMDGYGAVYTRSSVLPNSIPELE 1292
 DB 1552 -----IEEDKVEDIDEDIDEDIGEDKQDEVID-----LIYOKE 1583
 QY 1293 KAKSIGRDFSTFIPTDNTCGEPALVERIEKTYEGEIRIYKEVE-----IEEVE 1345
 DB 1584 KRIEYV-----KAKKKLEKKVEEGVGLKHHVDEMKYVQKIDIEVD 1626
 QY 1346 KEVEKY--GRTEMTLFORLAEGFNEIKODE-----NFVRELSKEMEFLDEIKMEASE 1397
 DB 1627 KEYSKALESKNDVT-----NVLKQNDFFSKVKNFVYVPAAPFISAAVAAASY 1677
 QY 1398 VEKLF 1402
 DB 1678 VVGFF 1682

RESULT 5
 R72826
 ID R72826 standard; Protein; 2482 AA.
 AC R72826;
 DT 27-FEB-1996 (first entry)
 DE Human mitosis.
 KW Cell cycle; M phase; mitosis; retinoblastome; mitosis; cell growth;
 OS Homo sapiens.
 FH Key
 FT region location/Qualifiers
 FT 1480..1659
 FT /label= internal_repeat
 FT 1660..1839
 FT /label= internal_repeat
 PN WO9511309-A2.
 PD 27-APR-1995.
 PF 24-OCT-1994; U12162.
 PR 22-OCT-1993; US-141239.
 PA (TEXA) UNIT TEXAS SYSTEM.
 PI Lee W, Zhu X;
 DR WPI; 95-170229/22.
 N-PSDB: 086851.
 PT Purified mammalian protein mitosis and agents that bind it and
 inhibit its action - used to promote cell growth or to inhibit cell

PT division and/or proliferation.
 PS Claim 4; Fig 8b; 61pp; English.
 CC R72829 is human mitotin. Mitotin is involved in the regulation of
 CC the mammalian mitotic cell cycle. Mitotin as with E2P-1 (see R72824)
 CC interacts with the retinoblastoma protein (the retinoblastoma tumour
 CC suppressor gene product). Mitotin is first synthesized at the G1/S
 CC boundary, it is then phosphorylated from S through M phase, and during
 CC mitosis, is closely associated with the centromeres/kinetochores at the
 CC mitotic spindle poles. Mitotin is necessary for a eukaryotic cell to
 CC enter the M phase of the mitotic cell cycle and its degradation is
 CC necessary for a cell to advance on to the next stage. Mitotin is thus
 CC useful for controlling cell growth as overexpression of mitotin prevents
 CC a cell from exiting the M phase.
 CC An anti-mitotin antibody, antibody fragment or a phosphorylated mitotin
 CC mutetin (or nucleic acid encoding it) can also be used to inhibit cell
 CC division which is particularly useful for the study of the cell cycle.
 CC A further use is to control hyperproliferative cells, and so control
 CC diseases such as psoriasis and breast cancer. It can also be used to
 CC block gametogenesis of an immature gamete.
 SO Sequence 2482 AA;

Query Match 2.0%; Score 147; DB 1; Length 2482;
 Best Local Similarity 18.4%; Pred. No. 0.024;
 Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 LVERLEKAESEKLLLEAVIEEVE-----KEVE-KVADTEMTLFLORLEGEFKELOQ 432
 DB 315 LSETLSEKEMSSIIISLNKREIELLOENGTLKINASLNQEKMLIÖK-SEFANYID 373
 QY 433 DEENVRELS-----KEEKEL----- 450
 DB 374 ESEKESISELDQYKOEKLLILORCETGNAYEDLSÖKTKAAQEKSKLECLINESTICE 433
 QY 450 ---NEIÖMATEVEKLFGRALP-IRKULMALAPSNFLANHETIKYVGSKLPGH----- 501
 DB 434 NKNKNE-----ÖLKEFAKHEQHEFLTKLAFABERNÖNMLELEYÖQÖLRSMTDNÖNNS 489
 QY 501 KPEFSGWEDYFSIYVAKICSSRRIPRYFRKSPRI-----CCGDSRQÖLÖFSHGK 551
 DB 490 KSEAGGLKÖEIMTLKEÖKÖKÖEYNDLLÖENÖLÖKVKYTKHEÖQÖNDESEPIR----- 544
 QY 552 HNLSPAHSHNÖVPPKNSGCKFPKVDALMWKMGÖFAKTAIVAFILISVASKADAVDAL 611
 DB 544 -----NSYKERESERKÖNÖ-CNFKPÖNDLEVKETISLDSYNAÖLVÖEAMIRKNEKLÖÖSE 596
 QY 612 KTCTCP-----LKECRLEL-AKCIÖSNP-----ACAN 637
 DB 597 KKEKCIÖHELOIRIGDLENSNÖMÖSÖEISGLKÖCEIDAEKYSISGPHELSTÖQNDNAH 656
 QY 638 VAC-IÖTCNNRDETE--CÖIKGDLFENSVDDEFNECAVSKKCV--PRK--SÖVGDF- 690
 DB 657 LÖCSLÖTÖMKNLEIEKIEIÖÖAEKYE--LVTEIÖND--SRSECIÖTARKEAEVÖGKLL 711
 QY 690 ----PVPDSVYÖKFDMDÖSGKMFITRGANP-----TFDAFC 725
 DB 712 NEVKILNDSGLHGLVDEIDPÖGEF--GEÖPNÖCHPVSÖLAPLDESNEYHLLISÖKEY 768
 QY 726 ÖLHEFTEENKILVNLÖWRIRTPDÖGFÖTÖRSVÖKÖFVÖPKYPGILYHNDEYLLYÖD-- 784
 DB 769 ÖMÖFALÖÖK-----FLSÖSEHKILHÖHÖQ--MSKMSÖLÖQYVDSL 810
 QY 784 --DWYLLSKVENSPEYDIFVYKGRND-----AMÖGYGSVLYÖRSAYL 826
 DB 811 KANLVLSTLÖRNFÖGDLVKEMÖGLEGLVPSLSSÖVDPSSLSLDSSEFY--RALL 868
 QY 827 PE-----SIPELÖTÖAÖ--KVGRDÖNTÖFKIDNTCÖGPEPL-----VERLEKÖVEÖE 873
 DB 869 EÖQÖDMSLNLNÖGAVSANCÖSVDÖFCSÖLÖENLÖTRKÖTÖSAPKÖYÖELLESCEYÖR 928
 QY 874 RITIKÖVEELIEEVEÖK--VRÖKEVTLFSKLPÖG-----FKELÖRDÖENFLRÖLSÖE 922
 DB 929 ÖSL-----EKÖEÖKMSÖGÖIKNÖKIEIÖELBÖLLSSÖRÖELDCLKÖYLSÖNEGÖQÖKLTÖSV 984

QY 923 EMDVLDGLKMEATEVEKLF-----GRALIRKLMÖVATÖH-----CETS-- 961
 DB 985 TLEMESKLAÖKÖTÖLÖSLÖEVARLÖLÖGÖDLSRSLÖGDTEDALÖGRNESÖDISÖE 1044
 QY 961 -----PCHDRIRFFSSDÖIGRLÖITRRIRINGTÖLLKILPI-----ÖSADL--- 1003
 DB 1045 HÖTSETTERIPÖHD-VHÖIQDKDÖÖDLNÖDIERIKÖTÖGALK--PTGÖSGÖSPDÖTYÖE 1100
 QY 1003 -----RTGGRS-----SRPLSAFRSÖFSGÖIDYIPLPSÖK----- 1034
 DB 1101 PÖGÖEDTÖGÖSÖCISÖLSFSGÖNALVPMÖDÖLÖNÖDÖIHNLÖÖRÖVÖTÖSÖNÖRLLHÖVED 1160
 QY 1034 -----NÖKELTAPLÖLLKVLGVLACÖFLVPSADÖVADLÖKTÖACLLKÖGRIÖELAKC 1084
 DB 1161 RÖDKVÖSLNÖKÖELDSÖKHLÖEVÖLMT-----KÖEAC-----IÖLEKI 1199
 QY 1085 IANPACÖANVACIÖTCNNRPDETECÖIKGDLFENSVDDEFNECAVSKKCVPRKSDÖG- 1144
 DB 1200 V-----GEL-----KÖ-----ENSDLSE 1212
 QY 1144 --EPAPDSVÖVÖNÖNISDÖFNKÖWYITSGLNPTÖPÖFCÖLHEFTE-----GÖDKKLVGN 1197
 DB 1213 KLEYFSDHÖELLÖRVÖTSE-----GLNS-----DLEMÖADKSSREDIGÖVÖAKVN 1258
 QY 1198 ISMRITLDSGÖFTÖRSVÖKÖFVÖDPNÖGVLVNHÖNEYH-----YÖDDWYLLSSÖKIÖENK 1252
 DB 1259 DSKMRFLD-----VENÖLSRITSEKASIEHEALYÖLEADÖLEVÖQTÖKLENDNEK 1310
 QY 1253 P-----EDYIFYYÖGRANDÖMDÖYGÖAVYÖTÖRSSVPSNIIPELEKAKSIGÖRÖSTFI 1306
 DB 1311 ÖKVYVCEELSVTSEBRNÖL--RGELDTMSKKTÖLALDÖLSEKMEKÖTÖELESÖSECL 1367
 QY 1307 RTDNTCÖPÖPALVÖRIÖEKYÖEÖEÖRIÖVÖEÖIEÖE--VEKÖYÖKVÖRTÖMTLÖFORLÖ 1363
 DB 1368 HÖIÖVÖAEVÖKÖETÖLLÖTÖLSDVSÖELDKHÖLÖKÖLÖSÖLEDSÖALSÖLTÖCELENÖTÖ 1427
 QY 1364 EGFNELKÖDEÖNFVÖR-----LSKEMFÖLDEIK-MEASÖYÖK 1400
 DB 1428 ----ÖLNKÖKELVÖKESÖIÖANLSESÖDYÖKUNVÖSÖALÖRÖALVÖK 1468

RESULT 6
 W23996
 ID W23996 standard; Protein; 2482 AA.
 AC W23996;
 DT 28-MAY-1998 (first entry)
 DE Human mitotin amino acid sequence.
 KW Mitotin; phosphoprotein; mitotic cell cycle; antibody; analogue;
 KW inhibition; M phase; Antagonist; hyperproliferative cell; cancer;
 KW leukemia; lymphoma; chromosome segregation.
 OS Homo sapiens.
 FH Key
 FH Location/Qualifiers
 FT 258..280
 FT /note="leucine heptad repeat"
 FT Domain
 FT 340..362
 FT /note="leucine heptad repeat"
 FT Domain
 FT 564..593
 FT Domain
 FT 1387..1443
 FT Domain
 FT 1885..1962
 FT Domain
 FT 2146..2188
 FT Domain
 FT 2165..2187
 FT /note="leucine heptad repeat"
 FT Misc-difference 2188
 FT Misc-difference 2300
 FT /label="Bipartite targeting motif"
 FT /note="Optionally C or G"
 FT Misc-difference 2189
 FT Misc-difference 2301
 FT /label="Bipartite targeting motif"
 FT Misc-difference 2303
 FT /note="Optionally A or T"
 PN US5710022-A.
 PD 20-JAN-1998.

PF 24-OCT-1994; 328254.
 PR 24-OCT-1994; US-328254.
 PA 22-OCT-1993; US-141239.
 (TEXA) UNIV TEXAS SYSTEM.
 PI Lee W, Zhu X:
 DR MPR: 98-109817/10.
 N-PSDB: V09076.
 PR New isolated mitosis protein and gene - useful for, e.g. developing
 PT products for therapy and diagnosis of hyper-proliferative disorders
 PS such as cancers or psoriasis
 PS Claim 1: Column 40-52: 43pp. English.
 CC This is the amino acid sequence for mitosis, a phosphoprotein
 CC necessary for the cell to enter mitosis. The protein's degradation is
 CC also necessary for the cell to advance into the next stages of mitosis.
 CC The mitosis protein, can be used to control the growth of cells. An
 CC anti-mitosis antibody, a mutant or a non-functional analogue of mitosis
 CC can inhibit the mitotic cell cycle by preventing the cells from entering
 CC the M phase, and over expression of mitosis or its functional
 CC equivalent, would inhibit the cycle by preventing cells from leaving the
 CC M phase. Antagonists to this protein can be used to control
 CC hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease,
 CC psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast
 CC cancer, sarcomas and other neoplasms, bladder cancer, colon cancer,
 CC lung cancer and various leukaemias and lymphomas). Reintroduction or
 CC supplementation of lost mitosis function by introduction of the protein
 CC or nucleic acid encoding the protein into a cell can restore defective
 CC chromosome segregation, which is a marker of progressing malignancy.
 CC Malignant proliferation of cells can then be halted. The protein
 CC can also be used for the detection and diagnosis of hyperproliferative
 CC cells.
 SO Sequence 2482 AA;

Query Match 2.0%; Score 147; DB 1; Length 2482;
 Best Local Similarity 18.4%; Pred. No. 0.024;
 Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 LVERLEKAESEKLLKEAVIEIEVEV---KEVE-KVBDTEMLYFQRLLEGSKELQ 432
 DB 315 LSETLSLEKKESSILSLNKRREIELTQENGTLKEINSLNQEKMLIQK-SESFANYID 373
 QY 433 DEENFRELSS---KEKEEIL----- 450
 DB 374 ERENSISLSOYKQKILLQRCBETGNATEDLSQYKAAQENSKLECLINECTSLCE 433
 QY 450 ---NELQMEATEVEKLFGRALP-IRKLNALAPHSNFLANHETIKYVYQSKLPGR- 501
 DB 434 NRKNELE---QLKEAFKEHQEFLTKLAFABERNONMLLELVQALISEMINDQNNNS 489
 QY 501 KRFGMGWDYGSIVYAKICSSRIIPRFKSPRI-----CGGLDSRGLOLJESHGK 551
 DB 490 KSEAGGLQELMTLKEQDNKKQKEVDNLLQENQMLVMKTKHDCQMLSEPIRI----- 544
 QY 552 HNLSPASINONVPGKSGCKFPKDVALLMWEKVGQFAKTAIVAFILTSVAASKADAVDAL 611
 DB 544 ---NSVKRESEBERNG-CMFKPQMDLEVKELISLDSYNAQVQLEMLNKKELKLOESE 556
 QY 612 KTCICL-----IKERLEU-AKCSNP-----ACAAN 637
 DB 597 KEKEDLHELOTIRGDELSNLDQMSQELSGKDCEDIAEKYISGFHELSTQNDNAH 656
 QY 638 VAC-LQFCNNRPDETE--COIKCGDLFENSVDFNCAVSRKKCV--PRK--SDVGF- 650
 DB 657 LQGLSLQTTMKNLKELEICETLQAEKYE--LVTELDN--SRSECIATRKMAEVEKEL 711
 QY 690 ---PVDPSPVLYQKFMKDFSGKMFTRGLNP-----TFAPFC 725
 DB 712 NEVKIINDSGELHGLVEDIPGEF---GEQPNQHPVSLAPLDESNSYEHILTSDKEV 768
 QY 726 QLHEFHTEENKLVGNLSWRIRTPDGGFTTSAAVQKQVQDPKRYGILYNHDEYLLYQD-- 784
 DB 769 QMHPAELOEK-----FLSLQSEHKILHDQHC--MSKSKSELTQTVDSI 810

QY 784 --DMYILSSKVENSPEDYIFVYKGNND-----ANDGCGSVLYTRSAVL 826
 DB 811 KAEVLSTVLRNFQGLDKENQMLGLEBLVPSLSSCVPDSSSLSDSSFY--RALL 868
 QY 827 PE-----SIPELQTAQ-----KVRGFNFPIKTDNMGPEPL-----VERLEKVEGE 873
 DB 869 EQTGMSTLSNIEGAVSANQCSVDEFCSSLOEBNITRETPSARAKVEELISEVYR 928
 QY 874 RTIKVEVELEVEVEK---VRDKEVTLFSKLEFEG-----FKELQDENFLELSKE 922
 DB 929 QSL-----EKLEKMEQGIKNNKEIOLEIOLLSEROEIDLRKQYLSNEQMOQKTSV 984
 QY 923 EMDVLGKMEATEVEKLP-----GHALPIRKMANATP-----CFTS-- 961
 DB 985 TLEMEKSLAEKKQOTQSLLEFVARLOGLDLSRSSLLGDTDAIQGNESCDISKE 1044
 QY 961 -----PCHDRIRFFSSDDGIGRLGTRKRINGTEFLKILPPI-----QSADL--- 1003
 DB 1045 HTSETTERPKHD-VQICDKDAQDLDNIDIKETGTALK---PTGCSGQSPDTNVE 1100
 QY 1003 -----RTGGRS-----SRLSAFRSQSGKGTIDYPLPSK----- 1034
 DB 1101 PGEEDTQSSSECISELSFSGNALVPMDFLGNQDINHNLQIRKETSNNMLRLHVIED 1160
 QY 1034 -----NEKELTAPLLKTLGVLCACAFIVPSADAVALKACALLGCRTELAKC 1084
 DB 1161 RDKVSLNLEKKEKLSKHLQEOVLMT-----KTEAC-----TELEKI 1199
 QY 1085 IANPACANVACLQTCNNRPDETECOIKCGDLFENSVDDEFNECAVSRKKCVPRKSDIG- 1144
 DB 1200 V-----GEL-----KK-----ENSDISE 1212
 QY 1144 --EFPADPSVLYQNFNNDENGKWTITSGNPTPTDAPDCQHERHTE---GNKLVGN 1197
 DB 1213 KLEPYSCHQELLQVETSE-----GLNS-----DLEMHADKSSREDIGNAVAVN 1258
 QY 1198 ISMRKTLDGFFTSAAVQKQVODPNQGVLYNHDEYLAH-----YQDWTILSKIEKN 1252
 DB 1259 DSMKERFLD-----VENELSRIRSEKASIEHVALYLEADLEVQTEKLEKNDENK 1310
 QY 1253 P-----EDYIIVYRRGRNDAMDYGAVVYTRSSVLPNSIIPLEKRAAKSIGRDFSTFI 1306
 DB 1311 QRVIVCLEEELSVTSSEKNQL---RGELDTMSKRTTALDQSEMKKEKTEQLJESHQSBCL 1367
 QY 1307 RTDNTGPEPALVERIEKTEVEGERIIVKEVEELIEE---VEKEVEKVGREMLLPQRLA 1363
 DB 1368 HGIQVAEVEKKTLLDTLSSDVSSELKDKTHQEKQLQSEKDSQALSTKCELENOIA 1427
 QY 1364 EGFNELKQDEENFVRE-----LSKEMEFLDEIK-MEASEVER 1400
 DB 1428 ---QLNKEKELLYKESSESQARLESSEDEKLVNSKALEALVEX 1468

RESULT 7
 ID W71295 standard; Protein; 1312 AA.
 AC W71295;
 DI 25-NOV-1998 (first entry)
 DE Human homologue of yeast RAD50.
 KW Human homologue; yeast RAD50; Drosophila Septin-2; Acyl-CoA synthetase;
 OS Homo sapiens.
 PN WO9838306-A1.
 PD 03-SEP-1998.
 PF 27-FEB-1997; WO03159.
 PR 27-FEB-1997; WO-003159.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Dolganov G;
 DR WPI: 98-481207/41.
 DR N-PSDB: V59979.
 PT Novel human immunomodulatory poly(peptide(s) - have homology to the
 PT yeast RAD50 or Drosophila Septin-2 proteins
 PS Disclosure; Pages 136-140; 155pp. English.

The present sequence represents a human homologue of the yeast 5. cerevisiae gene RAD50. The present sequence has 35% overall homology to the yeast RAD50 gene, and is expressed in activated T-cells, testis, fetal liver and heart tissues. The specification CC also describes sequences encoding human homologues of the yeast RAD50, the Drosophila Septin-2 and Acyl-CoA synthetase. The CC proteins have immunomodulatory activity. The nucleic acids and proteins can be used to identify activated T-cells in a sample population. They can also be used to isolate and identify sequences encoding other proteins or other compounds having immunomodulatory activity.

CC Sequence 1312 AA;

Query Match 1.8%; Score 135.5; DB 1; Length 1312;
Best Local Similarity 19.7%; Pred. No. 0.073;
Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

```

382 RLEKTAEEGE---KLIIKAEVEIEEVEKEVEKVRPTMTLFGRLLEGFELOODENF 437
398 KLVREHGEAEATANOAMDFAKETLKOKOIDEIRD-KKTGLRIIEKSEI----- 450
438 VRELSEKEEILNELOMEATEVEKLFGRALPIRKLMALAPHSNFL-----ANHETIKY 491
450 ---LSKKO---NELKNVYELQOLEGSSDRILEDOELIAERELSAEKNSVETLKM 502
492 YGSKLPGRHKRFESWGEDYFGSIIVAKICSSRIPIRYFRKSPRICCGLDNSGLOFSHGK 551
503 EVIS-LQNEK-----ADLDRIIRK-----LDQEMEQ-LNH-- 531
552 HNLSPAHSINQVPRKNSGCKEPPKDVALLMWEKMGOFKATAIATFILLSVASKADAVDAL 611
531 -----HTTRTROMEMILTKDRAKDKQIRKIKSRHSDELTSILGYF-----PNKQLEBWL 580
612 KTCCLKEKELFLACISNPACAAVACLOTCNNRPDETECOIR-----CGDLFE 662
581 HSKSKEINQTRDLAKL--NKLASSEQNKNNHINELKRREQLSSEYEDKLEVDGSGDF 638
663 NSVYDEFNECAVSRKCKVPRKSDVGFPPDPVSVLYOKTDMKDFSGKWFITRGLNPTPDA 722
639 ESDIDRLKE-----EIEKSSKQRAMLAGATAVVSQ 668
723 FDCQLHEFTEENKLVGNLSWRIRTPDGGFFTRSAVOKFVOPKPYGILLYNDNEYLYXQ 782
669 FITOL-----IDENOSCCPVOCORV-----FOTAEILOEVIISD----- 701
783 DDWYILSKVNSPEDYIFV--YKGRNDAMDYGGSVLYTRSAV-LPESIIPELOTAQ 839
701 ---LQSKLAPDKLKTSELSKKRERDEMGLVPRQSIIDLKEKEIPELNNKLQ 755
840 KYGRDFTFIKTD-----NTGCPPEPLVLELEKVEEGERTIIKE-- 880
756 NVNRDIQR-LKNDIEEQETLLIGTIPPEESAKVCLDVTIMERQOMLKVYERKIQQA 814
880 -----VEEIEEVEKVRDEKVLFLSKLFEGRKELODEENF-----LRELSKEE 923
815 KLQGLIDRLTVQVQNOEKQKHLDVYSKTEINRKLIDQOQOIOHLKSTNELKSEK 874
924 MDVLDGLKME-----ATEVEKLF 941
875 LOISTNLGRQQLLEQVLELSTEVOSLY 902

```

RESULT 8
W22775
ID W22775 standard; Protein: 1312 AA.
AC W22775;
DT 21-DEC-1998 (first entry)
DE Human RAD50.
KW Human; RAD50: DNA repair; tumour suppression; cancer; Septin-2;
OS Homo sapiens.
PN W09727284-AZ.

PD 31-JUL-1997. 001299.
PF 24-JAN-1997. 001299.
PR 17-JUL-1996. US-687080.
PR 26-JAN-1996. US-562126.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Dolgavov G;
DR N-PSDB: T75237.
DR N-PSDB: T75237.
PT Human tumour suppressor gene RAD50 - useful to detect
PT predisposition to, decrease risk of and treat cancer, also Septin-2
PT homologues
PS Claim 5; Page 82-86; 195pp; English.
CC The human RAD50 (hRAD50) is involved in DNA repair and has tumour
CC suppression activity, can be used to detect predisposition to, decrease
CC the risk of or treat cancers, e.g. acute myeloid leukaemia,
CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,
CC therapy related acute myeloid leukaemia, refractory anaemia or refractory
CC anaemia with excess blasts. Also disclosed in this invention is Human
CC Septin-2 homologues of which may be used as targets for cancer therapies
CC and central nervous system directed treatment methods, and to measure the
CC proliferative potential of selected cell types.

CC Sequence 1312 AA;

Query Match 1.8%; Score 135.5; DB 1; Length 1312;
Best Local Similarity 19.7%; Pred. No. 0.073;
Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

```

382 RLEKTAEEGE---KLIIKAEVEIEEVEKEVEKVRPTMTLFGRLLEGFELOODENF 437
398 KLVREHGEAEATANOAMDFAKETLKOKOIDEIRD-KKTGLRIIEKSEI----- 450
438 VRELSEKEEILNELOMEATEVEKLFGRALPIRKLMALAPHSNFL-----ANHETIKY 491
450 ---LSKKO---NELKNVYELQOLEGSSDRILEDOELIAERELSAEKNSVETLKM 502
492 YGSKLPGRHKRFESWGEDYFGSIIVAKICSSRIPIRYFRKSPRICCGLDNSGLOFSHGK 551
503 EVIS-LQNEK-----ADLDRIIRK-----LDQEMEQ-LNH-- 531
552 HNLSPAHSINQVPRKNSGCKEPPKDVALLMWEKMGOFKATAIATFILLSVASKADAVDAL 611
531 -----HTTRTROMEMILTKDRAKDKQIRKIKSRHSDELTSILGYF-----PNKQLEBWL 580
612 KTCCLKEKELFLACISNPACAAVACLOTCNNRPDETECOIR-----CGDLFE 662
581 HSKSKEINQTRDLAKL--NKLASSEQNKNNHINELKRREQLSSEYEDKLEVDGSGDF 638
663 NSVYDEFNECAVSRKCKVPRKSDVGFPPDPVSVLYOKTDMKDFSGKWFITRGLNPTPDA 722
639 ESDIDRLKE-----EIEKSSKQRAMLAGATAVVSQ 668
723 FDCQLHEFTEENKLVGNLSWRIRTPDGGFFTRSAVOKFVOPKPYGILLYNDNEYLYXQ 782
669 FITOL-----IDENOSCCPVOCORV-----FOTAEILOEVIISD----- 701
783 DDWYILSKVNSPEDYIFV--YKGRNDAMDYGGSVLYTRSAV-LPESIIPELOTAQ 839
701 ---LQSKLAPDKLKTSELSKKRERDEMGLVPRQSIIDLKEKEIPELNNKLQ 755
840 KYGRDFTFIKTD-----NTGCPPEPLVLELEKVEEGERTIIKE-- 880
756 NVNRDIQR-LKNDIEEQETLLIGTIPPEESAKVCLDVTIMERQOMLKVYERKIQQA 814
880 -----VEEIEEVEKVRDEKVLFLSKLFEGRKELODEENF-----LRELSKEE 923
815 KLQGLIDRLTVQVQNOEKQKHLDVYSKTEINRKLIDQOQOIOHLKSTNELKSEK 874
924 MDVLDGLKME-----ATEVEKLF 941
875 LOISTNLGRQQLLEQVLELSTEVOSLY 902

```

RESULT 9

W85011 standard; Protein; 968 AA.

AC W85011: (first entry)

DE 08-FEB-1999 (first entry)

DE p85alpha-green fluorescent protein fusion product.

KW Human; p85alpha gene; fusion protein; green fluorescent protein; GFP;

OS Intracellular signalling; chimera.

OS Chimeric - Aequorea victoria.

PN M09845704-A2.

PD 15-OCT-1998.

PE 07-APR-1998; DK0145.

PR 07-APR-1997; DK-000392.

PA (NOVO) NOVO-NORDISK AS.

PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,

PI Tullin S;

DR WPI: 98-594491/50.

DR N-PSDB: V71026.

PT Determining effect on signalling pathways in live cells from

PT redistribution of luminophores - specifically fusions of green

PT fluorescent protein with a signalling component, and new apparatus,

PT particularly for identifying toxins and potential therapeutic agents

PS Example 15: Pages 96-98, 326pp; English.

CC The present sequence represents a human p85alpha-green fluorescent

CC protein fusion product. The fusion protein is used in an assay

CC that exemplifies the invention. The specification describes how

CC quantitative information about the influence of a molecule on a cellular

CC response is obtained by recording the variation, caused by the molecule,

CC on mechanically intact living cells, in the spatially distributed light

CC emitted from a luminophore present in the cells. The variation in light

CC emission is processed to provide information that correlates spatial

CC distribution to the degree of the molecule. The method is used to

CC identify agents that (in)directly affect intracellular signalling,

CC especially to screen for potential therapeutic agents or toxins, and

CC to identify new drug targets.

SQ Sequence 968 AA;

Query Match 1.8%; Score 135; DB 1; Length 968;

Best Local Similarity 19.8%; Pred. No. 0.05;

Matches 180; Conservative 111; Mismatches 298; Indels 318; Gaps 49;

QY 687 GDFPVPDSVL-----VQKED-----MKDFSGKFLTRGILPTFDAPDQLEHFTTEEN 735

DB 52 GKLVPWPLVTLTLYGOCSTRYDHHKQ-----HDFKSGAMEGYVOER 97

QY 736 KLVGMLSRIRTPGCGFTBSAVQKRVQDP-----KYPGILVHNDNEYLTYOD 783

DB 98 TI-----FKRDOGNYKTRAEV-KFEGDGLVNRLEIKGIDFKEDGILGHKLEIYVNSH 149

QY 784 DWYILSKVENSPEYIFVYKKGNDAMDGGSVL-----YTSNAVLPSE--IIPELQTA 837

DB 150 NYIIMADKQKNG-----IKVNFKIRHNIED--GSVQADHYQONTPLGDPVLLPD----- 199

QY 838 AQXGROENTIKINDNTCGPEPLVERLEKKEVEGERTIIEVEETIEEVEKVRDKEVTL 897

DB 199 -----NHLSTQSLASDNP-----EKRDHMLV-----LEETAGAITL 232

QY 898 FSKLEFGEKLEORDEENFLRELSEKEMD---VLDGLKMEYVEKLFPGALPRTIKIMAVA 954

DB 233 -----GMDELYK-----SGLSMSAEGYQYRALDYKREKEDIDLGLITLVNKGSLVA 283

QY 955 THCTSCPCRIIRFFSSDDGIGRLGTRKRINGTFL-----LKIIPQISADLRTTGGGS 1009

DB 284 LG-FSDGQEARPEIIGLWNGYNETTERGDPCTVEYVIGRKKISFP-----TPKFRP 335

QY 1010 SRPLSAFSGSGKIFDI-----VPLPSKNEL-----KELTLPILKLVGV-----IACAPLI 1057

DB 336 PRPLPV-APGSSSTEADVEQALITLPDLAEQFAPIADIPILIKLVEAIEKKGLECSTLY 394

QY 1058 -----VPSAD-----AVDALTKTACILKCKRIE----- 1081

DB 395 RTOSSNLAELRLDCCDPSVDLEMDIVHVLADAFKRYLLDLPNPVIPAAYSEMISLA 454

QY 1081 -----LAKCIANPAC-----AAVACLQICNNR----- 1104

DB 455 PEVOSSEYIQLKILRSPISIHQYWLLOYLKHFELSTQSSKNLINAVALSEIFSP 514

QY 1104 -----PDEEQIKGCDLFENSVDENCAVSRKCVPRKSDGEPAPDPSVLY 1154

DB 515 MLRFSAASSDNTENLIK--VIELISTENNERQPA--PALPK-----PKPTIVA 562

QY 1155 -----ONFNISDPNGKMY-----ITSGLPPTFA-----FDOOLHEFH-----EGDN 1192

DB 563 NNGMNNMNSIQNAEYWGDISREVENKTRJADGTFLVRASTIMHGDYTLTRKGGNN 622

QY 1193 KLVGNISWIKTLIDS--GT---FTRSVQKRVQD--PNOGVLYNH--DNEYLH----YOD 1240

DB 623 KLI-----KIFHRDGKYGFSPLTFSSVVELINHYRNSLAQNPBKIDVLYVSKYQ 677

QY 1241 DWYILSKIE--NRPEDYIFVYGRNDAMDGGYGAIVYTRSSVLPNSIIPELERAKS 1297

DB 678 DQVVKEDNIEAVGKLUHE-----NT---QFOEKSRE 706

QY 1298 IGRDESTFRTDNTGCPPEALVERIEKTV-----DEGERIYKEVEEIEE-VEKEVE 1349

DB 707 YDRLYEYTRISQLOMKRTRIAEAFNETIKIFEBOCQTOEYSREYIEKFKRGNEKEIO 766

QY 1350 KVGRTMELFQRLAGFVELKODEENFVRELKREMEFLDEIKMEASEVEKLFKALP-- 1408

DB 767 RI-----MNYDKLKRISIEIISDRRLLEDLKKOAEY-----REIDKRNNSIKPDL 814

QY 1408 --IRKVR 1412

DB 815 IQLRKTR 821

RESULT 10

W54145 standard; Protein; 1639 AA.

AC W54145: (first entry)

DE 23-SEP-1998 (first entry)

DE P. falciparum synthetic gp190 protein.

KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;

OS monoclonal antibody; passive immunisation; parasite.

OS Plasmodium falciparum.

OS Synthetic.

PN M09814583-A2.

PD 09-APR-1998.

PE 02-OCT-1997; E05441.

PR 02-OCT-1996; DE-040817.

PA (BUJA) BURARD H;

PI Bujard H, Pan W, Tolle R;

DR WPI: 98-240088/21.

DR N-PSDB: V21451, V35363.

PT Recombinant production of complete gp190/MSP-1 Plasmodium surface

PT protein - useful in anti-malaria vaccines, also stabilising genes by

PT reducing their AT content

PS Example 1; Fig 3c; 48pp; German.

CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1

CC (merozoite surface) protein. The gene encoding this protein has been

CC stabilised by reducing the AT content of the nucleotide sequence. Such a

CC protein is useful in vaccines against malaria or for producing monoclonal

CC antibodies (for passive immunisation). The complete gp190 protein can now

CC be produced outside the parasite and has, at least over extended regions,

CC the native pattern of folding. Larger amounts of the protein can be

CC produced recombinantly than would be possible using the parasites as

CC source.

SQ Sequence 1639 AA;

Query Match 1.8%; Score 134; DB 1; Length 1639;

Best Local Similarity 18.0%; Pred. No. 0.14;

Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

Thu Oct 14 07:57:24 1999

us-09-075-375-2.rag

Page 9

QY	184	VDOFNCAVSRKKCVPRKSDVGEFFVPDNRNAVONFNMDQSCSKYTITSGLNPIFTDADC	243
Db	249	IEININELIEBSKTTIDK-----NKNAITREEKKKLYAQYDLS-----LYNK	290
QY	244	OLHEFHENDKLVLNWRKTLTLDGFFTRSAVOTF--VODP-----DLPALYNHNE	295
Db	291	OLEBAH-----NLISVLEKRIDITLKNENIKELLOKIMKINPPRPAANGNPNTLIDKNK	346
QY	296	FLHYQDWTYLLSSQINPKPDY-----IFYVYRGNDAMDGYGGSYITRSPITPEJII	349
Db	347	IEEHKEIKETAKTIFKNIDSLFTDPLELEYLREKKNKID--ISAVETKESTEPNE-Y	403
QY	350	PNLOAKASVQDORNFNITTNSCGEPPLVRLKTAEEBKKLII-----KEAVIEEB	404
Db	404	PNGVYTPLYS-NDINNALNLSFG--DLINPDYETKPSKNITYDNERKKFINETKEK	459
QY	405	VEREYEXVRDEMTLFOLEGEFKELOODENFVRLSKEEKEILNEL-----OQEA	457
Db	460	IKIEKKLI-----ESDKSYEDSKSLINDITKEYEKLINELYDSKFNANNIDLT	507
QY	458	EVEKEFJGR--ALPIRLMALAPNSNLANHEITKYVYSKJPHGKRSWGEDY-FGSI	514
Db	508	NEKMGKRYKYVKEKL-----THHNTPASTENSKHL-EXLTKALKY--MEDYSLEKNI	558
QY	515	VVAKTSSRIPRYEPRKSPRJCCGIDSKGLOLFSHGKHNISPHASINQWVPNGSGC--	572
Db	559	VEREKL-----KYIK-----NLISKINEL--ETLVENIKDEBOLIEK	595
QY	572	-----KRPDVALAMWEMQOFAKTALVALFISVASKADAVADALKTCTCLIECRLE-	625
Db	596	KITDENKRPDEKILEV-----SDIVYQVQYKILM-----NKIDELKKTOLILKYNELKH	645
QY	625	-----LAKCISN-----PACAA-----NYACQOTONNPDPT	651
Db	646	NIHVPNSIKQENKOEPRYLLVLKKEIDKLKVMPEYSELINEEKKNKITQSDNSEPST	705
QY	652	ECOI-----KCGDLFENVSDNEFMECAVSRKKCVPRKSDVGEFPVDPDSVLQ--	700
Db	706	EGEITGATTPRGOQASALEG--DSVQAQOQOQOAP-----PYVYVPPAKAOP	756
QY	700	-----KFDKDFSGKMFITTRGLNFTPAFCOLHEHTENKVLNLSMRI	745
Db	757	TPPAVNNKTEVSKDYLE-----KLYEPL-----NTSYIC	788
QY	746	RPPDGFTRSAVQKFOVDPKPYGLILNHD--NEYLQDOWMYLISKRYEN-----SPED	798
Db	789	H-----KY--ILVSHSTMEKILIKQ--YKITKEESKLSGCDPLD	824
QY	799	YIF-----VYKGRNDAMDGYGGSV-----LYTSAYL-----PESIIPELOTAAQ	839
Db	825	LLEIYONNIPVMI-----SMFDLSNLSLSQLMPEIYEREMVCMYLKDKNDKIKNLIEAK	880
QY	840	KVGHDFNTIKTIDNTGPER-PLVRLKEKVEEGERT-----IIEVEEYI-----	884
Db	881	KV-----STSVKTLSSSMQPLSTLPDOKREYVSANDTSHSTNLNSLKLENTILSGKNK	936
QY	884	-----EEVEKYADKEVTLTFSKLEPGEFELORDDEENLRELSEKMDVDLGJK	931
Db	937	NIYQELIOKSENFYKILKDSOTFYNESFINFVKSKADIDINSINDSKRK-----K	989
QY	932	MEATEVEKLFGRALPIRKILMAVANTHOCFTSPCHDRIFESSDDGIGRLGITRKIRINGFLL	991
Db	990	LE-EDINKI-----KKTJOLSDLYNNKKKLEBLFPRKKTYGKYKMQOIKRLT--LL	1038
QY	992	KILPPIQSADLRTTGGRSSRP--LSAFRSGESKGJFDIPLPSNNELKELTAPLILKLY	1048
Db	1039	K-----BOLESKILNSLNNKHYLONFSVFENK-----KKEAF-----	1071
QY	1049	GVLACAFLLIVPSADAVDALKTACCLIKCCR--ELAKCIANPACAAVACLOTCONRPE	1106
Db	1071	-----IAEENTIENTKILKIKRYKRLVYNGESSPLKTSIESITODENYASL	1119
QY	1107	TEQIKGCDLFENSVDENECVSRKKCVPRKSDVGEFFVPDNRNAVONFNINDSFNGRW	1166

QY 908 ---LORDENFURELSEKEMDVLGKMEATEV---EKUGRALPI-----RKL 950
 Db 297 QAPIEPKDDIHLSEKELKRIQIDSSDLSTEEKEFKLLIDIRDSLESEKEL 356
 QY 951 MAVATHTCTSPCHDRIRFFSSDDIGRLGTRKRINFTFLKLIPQIAD---LRTG 1006
 Db 357 L-----NRIGVSSNP-----LSEK---EKELKLKLDIQPYDINRLDPTG 396
 QY 1007 GRSSRPL-----SAFRSGSKGIDIVLPSPKMLKELTAPLLKLKLVG 1049
 Db 397 GLIDPSINDVRKQYKRDIONDALLHOSIGSTLKNKYLYENNNINNTATL----- 451
 QY 1050 VLACAFIVPSADAVAKATCACLKGCRIELAKCIANPACANAYACIQTONNPDTEC 1109
 Db 451 -----GADLVST----- 459
 QY 1110 QIKGDLFENSVD--FNECAVSRKCVPRK---SDIGEPADPSVLVQNINSDENG 1164
 Db 459 -----DNTKINRGIENFEKKNFKYSISSNYMLVDINERPALDNRLKMRITQLSPDR 510
 QY 1165 KWTITSGLNPTPDACQHEHTEGDKLV--GNISMRKLTDSGFFTRSAVQKRVDP 1222
 Db 511 AGYLENG-----KILORIGLEIKDVO----- 534
 QY 1223 NQGYLVHNDNYLHYQDWYILSKTENKPEYIEFYRGRDADGIGAVYTR--S 1280
 Db 534 -----IKOSEKEYIRI-DAKVPRKSIDTKIG--AOLINQENKALGLPKYTKLIT 584
 QY 1281 SVLPNSIIPLEKAKSGRDSFIRDTWCGPEPALVERIEKVEEGR--RIYKEY-- 1338
 Db 585 FNVNHRVANSNIVESAYILNEMKNNTIOSD-----LHKVTNYLVGNGEFVFTDTL 636
 QY 1338 -----ELIEVEKVEKVGRTENTLFLQAEENELKOEENFVVE----- 1380
 Db 637 PNIAECTHODELIEYQVHSGLYPESRSILHGPSKCV-ELRDSGSEIHEFGHAVDY 695
 QY 1380 -----LSKEENE-----FLDEIKMEASV 1398
 Db 696 AGYLLDRNOSDLVTSKRFIDIFKEGSNTL 725

RESULT 12
 Y01632
 ID Y01632 standard; Protein: 2954 AA.
 AC Y01632:
 DT 22-JUN-1999 (first entry)
 DE Amino acid sequence of centromere-associated protein-E (CENP-E).
 KW CENP-E; centromere-associated protein-E; ATPase activity;
 KW plus end-directed microtubule motor activity; chromosome congression;
 KW microtubule binding activity; chromosome movement; mitosis;
 KW cell proliferation; tumor; metastasis; vascular malformation;
 KW inflammatory disease; immune disease; angiogenesis; hypertension;
 KW insecticide; fungal infection; selective herbicide; fungicide;
 KW xenopus sp.
 OS Xenopus plant growth regulator; activator; cancer cell marker.
 PN M09913061-A1.
 PD 18-MAR-1999.
 PF 10-SEP-1998; 019231.
 PR 11-SEP-1997; US-058645.
 PI (REGC) UNIV CALIFORNIA.
 DR Cleveland DM Goldstein LSB, Sakowicz R, Wood KW.
 DR MPI: 99-229233/19.
 DR N-PSDB: X26819.
 PT Centromere-associated protein-E and related nucleic acid
 PS Clam 5; Page 66-67; 77pp; English.
 CC The present sequence represents CENP-E (centromere-associated protein-E)
 CC of Xenopus. The protein has at least one of plus end-directed microtubule
 CC motor activity, ATPase (adenosine triphosphatase) activity and
 CC microtubule binding activity. CENP-E is the motor that powers chromosome
 CC movement toward microtubule plus ends and is essential for congression
 CC of chromosomes during mitosis. Modulators of CENP-E can thus control
 CC cell proliferation. Agents that modulate CENP-E activity are lead
 CC therapeutic, bioagricultural and diagnostic agents, e.g. for treatment

CC of unwanted cell proliferation (typical of many examples are tumors and
 CC metastases; vascular malfunction; inflammatory and immune diseases;
 CC angiogenesis; hypertension; restenosis; and fungal infections), also as
 CC plant protection agents (selective herbicides, fungicides and
 CC insecticides) and plant growth regulators or activators for improving
 CC cancer cells.
 CC CENP-E is also a diagnostic marker for dividing cells, including
 SQ Sequence 2954 AA.

Query Match 1.7%; Score 129.5; DB 1; Length 2954;
 Best Local Similarity 19.0%; Pred. No. 0.77;
 Matches 260; Conservative 190; Mismatches 441; Indels 479; Gaps 64;

QY 202 SDVGEFVPDRNAYVQNFMKFSGKWTITGLNPTDAPDCOLHEPHMENDKLVGNLTW 261
 Db 459 SDMPSPPEID-DVCTERSDEFDALSMDSNGIDAENV-----LASKVTH 502
 QY 262 RIKTLGDFTRSAVQTVQDDLPALYNHNDNEFLHYQDWYITSSQIENKPPDYIFYV 321
 Db 503 REKT-----SLHOSMIDFGQISDSVQFHDSS-----KENQLOLYPKD----- 540
 QY 322 YGRNDAMDGYGGSYIYRSPTEPESIIIPNLOCAASVGRDENFTTIDNSGCPPEPVE 381
 Db 540 -----SGDMACGRKASP-EKELTISLQOQLOSKREE-----KKELYQ 574
 QY 382 RLE-KTAESEKLLIKRA-----VEIEVEKVEKVRTENTLF----- 421
 Db 575 SPELKIALEFQLSVKARKNLEMTNSREHSINAEVQTVDEKRYA--KEMSVLGDGYN 631
 QY 421 -----QRL-----LEGFELOO--DEENFVRELSEK--EKEI 448
 Db 632 ASNDDLODSSVDCKRLSSHDECIERKALDQKLYDLEEFLENKRSKENDOKOSQDP 691
 QY 449 LNEQNL-ENAEVEKLEGRALPIRKLMALBPSNFLANHEITKYVGSKLPGHKRSWGW 507
 Db 692 MESIQCEALMEK-----ANALEELALMRDNFDNIILENETLKEIND----- 736
 QY 508 EDFGSIYVAKICSSRRIPRYFRKSPRICGDSRGQLTSHGKNLSPAST----- 561
 Db 736 -----LERSLEKNEGT--NEFEILEKTOHEHAQDIHEIGSLKLVL 775
 QY 561 -----NONVPRGNCGCFPRKDVALMWKMGOPAKTATVATPILTSVAKADAVAKATC 614
 Db 776 ENAEYMNQNL-----EDLETK 792
 QY 615 PCLTECELELAKCISNPACANAYACIQTONNPDTECOIK-----CGDL 660
 Db 793 TKLLEDEIQLAE-----LKRADMLQKKNANFDLSVMSGDSERKLEEI 836
 QY 661 FE--NSVVDENECVAVSRKCVPRKSDVDGFVPPDSVLVQKFDMDKDSGWFTRGILNP 718
 Db 837 FOLKOSLSDAEATVROAKQECSTLSE-----NLELKEKMEPTS-NMYNQK--EK 883
 QY 719 TFDAFDQOLHEFTTEENKLVNLSMRIRTPDGGFFTRSAVQKFOVDDPKYQICILYNDNEY 778
 Db 884 AASLFEKQLETESKNYKMEADL-----QKEIQS-----AFNEIN-Y 919
 QY 779 LLYQDWYITSSKVENSPEDYIFVYKGRNDAMGCVGSVLYTRSAVLPSIIPLEQTAA 838
 Db 920 LN-----GLLACKV--PRDL-----SRVELEKKYKSESKOL 949
 QY 839 OKVGRDFNFITDITWCGP-----PPIVERLEKVV-----EEBER--TIKEYE 881
 Db 950 EKALEEKN-LENEVTCLESEKFLPNEVECIKNQISAKSEIMLLQBSHSASTISQOE 1008
 QY 882 EI-EEVEKVDKVEITFSKFEGRFELQDEENFLRELSEKEMDVLQGL--KKEA-- 935
 Db 1009 IIMOSQSOI-----LQLDVETHQSVQOTEQYI-EKKKNHDLLEFYIRNKSSEADL 1063
 QY 935 -TEVEKLEGRALPIRKLMALVAHTCTSP-----CHDIRFSSDDIGRLGTRKRIN 986
 Db 1064 LREMENLKGIMSEVAKIADTKHELEETIRDKEOULLHEKKYFFQAMQITFPI----- 1116

QY 987 GTFLKILPPIISADLRTTGRSSRP-----LSAFRSGSGK1 1024
 Db 1116 -TPLSDSLPSKLYE-----GNSODPIEINDYHNLIALATERNNIMVCLERTNSLKEQV 1169
 QY 1025 EDI-VPUPS-----KNEKELIAPLLIKLY-----GYIACAFIIPSA 1061
 Db 1170 IDLNTQSLQASIEKSDLOKPKQDLLEGVYKLLLEMLIKHLTDSQSIEKLENTL 1229
 QY 1062 DAVDALKTACILKGCRIELAKCIANPACAANVACLOTNNRPDETECOIKGDLFEN- 1120
 Db 1230 EYTERKIQTLOEEMKNTTIERNE-----LQT-NPEDKAEHDSLKODLSNIE 1275
 QY 1120 SYVDENECVAVSRKCVPRKSDLGEPAPDPSVLVONFISDENGKWTITSGLPPTDAF 1179
 Db 1276 QSTETQDELRAQOELREQKO-----LYDSFRQQLLDCSVGIS--PNDHAY 1320
 QY 1180 DCO-----LHEFTGDNKLYG--NISWRIKTLDGFFTRSAVOKFVODPNO--PGVLYN 1230
 Db 1321 ANQKVSLEVNLSQSEMLRGERDELQTSCKALVSELELLRAHVSVGENLEITKILNG 1380
 QY 1231 HDNEVLYHODDWYILSSKIEKNRPEDYIFVYRGNDAMDGYGAVYTRSSVLPNS--II 1288
 Db 1381 LEKEILLKSESESVLKSMLENKED--NNKLKEQAEYSKENGQSLSEVFGSGSKLV 1436
 QY 1289 PELE-----KAAS-----IGRDEFTFRTDTCGPEPALVERIKTYEGEGRITVKEVE 1338
 Db 1437 DEIEVLAKQAKAEERLEIKRDYFELVQJANTNLVEGKLETPLOADHEE-----DSID 1490
 QY 1339 EIEBEVKEV--EKVGRTEMTLFORLAEGNELKODEENFVRELSEKEME 1386
 Db 1491 RRSEMEIKVLGERLERNQY-LLERLQOEKLELSNK-----LEILLQKEME 1534

RESULT 13

PS0777
 ID PS0777 standard: Protein: 1654 AA.
 AC PS0777;
 DT 30-SEP-1991 (first entry)
 DE Sequence of the p195 protein of Plasmodium falciparum.
 KW Malaria vaccine; epitope; antigen; immunogen.
 OS Plasmodium falciparum.
 PN EP-154454-A.
 PD 11-SEP-1985: 301173.
 PE 21-FEB-1985: 301173.
 PR 22-FEB-1984: GB-004692.
 PR 26-SEP-1984: GB-024340.
 PR 21-FEB-1985: GB-004429.
 PA (WELL) WELLCOME FOUNDATION LTD.
 PI Holder A. Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
 DR WPI: 85-224845/37.
 DR N-PSDB: N50530.
 PT Cloned DNA sequence encoding plasmodium falciparum protein -
 PT useful for expressing the protein for use in vaccines against
 PT malaria
 PS Claim 6: Fig 1: 51pp: English.
 CC The sequence encoding the p195 protein of Plasmodium falciparum
 CC (N50530) and a peptide comprising at least one of its epitopes
 CC (see PS0777) are claimed. Also claimed is a vaccine for inducing
 CC immunity to malaria comprising the novel peptide or p195 or a
 CC peptide comprising at least one epitope when derived from the new
 CC DNA sequence, together with a carrier.
 SQ Sequence 1654 AA;

Query Match 1.7%; Score 129; DB 1; Length 1654;
 Best Local Similarity 17.5%; Pred. NO. 0.34;
 Matches 238; Conservative 214; Mismatches 499; Indels 408; Gaps 67;

QY 184 VDFNFCANVRKCVPRKSDVGEFVPPDRNAVONFNKDKSGWYIISGLNPTFDADC 243
 Db 249 IENINELIESKTIK-----KNNAKEEKKKKLVQAQYDLS-----LYNK 290

QY 244 OLHEFEMENDKLYGNLTMRKLTLDGGEFFTRSAVQTE--VODP-----DLPALYNHDE 295
 Db 291 OLDEAH-----NLSVLEKRIDTLKKNKIKELDKINEIKNPPRANSGMTPTLLDKNKK 346
 QY 296 FLHYODDWYILSSQIENKPDY-----IFVYRGNDAMDGYGGSVITRSTLPESII 349
 Db 347 IEEHKEIYE:AKTIKIFNIDSLFTDPLELEYLREKKNID--ISAKVETKESTEPNE-Y 403
 QY 350 PLYQAAASVGRDNNFTTNSGPEPPLVERLEKAEGEKILI-----KEAVEIEE 404
 Db 404 PNGVTYPLSY-NDINNALNELNSFG--DLINPDYIKESKNIYIDNNEKRTINIKKK 459
 QY 405 VEKEVEKVDTEMTLFORLLEGFELOODEENFVRELSEKKEITLNL-----OMEAT 457
 Db 460 IKIEKKKI-----ESDKSYEDRSKSLNDITKEKELINEIYSKFNNDIDL 507
 QY 458 EYEKLEGR--ALPIKLMALAPSNFLANHETIKYVGSKLPHQKRFSGWEDY-FGSI 514
 Db 508 NEKMGKRYEYKVEKL-----THNPTFASYSKHNH-EKLTALKY--MEDYSLRNI 558
 QY 515 VYAKICSSRIPIRPRKSPRIQCGLSRGQLFSGKHNLSPAHISINQNPKNSSGC-- 572
 Db 559 VVEKEL-----KYYK-----NLISKIENEI--ETIVENIKKDEQLFEK 595
 QY 572 -----KPRDYALMWWEKMGQFAKTAIYALFILSVASKADAVALTCTCLKECHLE- 625
 Db 596 KIKDKENKDEKILEV--SDIVKVOYQVILM-----NKIDELKKTOLILKNVELKH 645
 QY 625 -----LACISN-----PACAA-----ANACLOTNNRPDET 651
 Db 646 NIHPVNSYQENKQOEYUILLKKEIDKLYMPRVESLNEEKNKITGQOSNSPST 705
 QY 652 ECOI-----KCGDLFENSVDDEFNECAVSRKCVPRKSDVGEFVPPDPSVLVQ-- 700
 Db 706 EGETIGQAITTKPQOAGSALG--DSVQAQOQEQKQAP-----PYVPVPPAKAQP 756
 QY 700 -----KFDKDESGKWFITRGILNPTFDADCOLHEFTGDNKLYGNLSMRIRTP 748
 Db 757 TPAPAVNNKTENYSKLDYIEKLY--QFLNTSYCHRYILVSHSTNNERILQYK----- 809
 QY 749 DGGFTFSAVQKPVQ-DPYPGILYHNDNEVLYHODDWYILSSKIEKNRPEDYIFVYGR 807
 Db 809 -----TIKEESKUSCDPL--DLFNQNNITV--MTSMDSILYIHNHLMWYEKE- 858
 QY 808 NDAMDGYGGSVL--YTRSAVL-----PESIIPELQIAQKVGDEMTFIKTDTNCPGE 858
 Db 858 -----IGVFTILMEIYEKEVCNLYLKDNDKIKNLLEBAKRV--STSVKTLSSSSMQ 909
 QY 859 P-PLVERLEKKEVEGET-----IIRKEVEI-----EEVEKV 890
 Db 910 PLSLTPDKREYVANDOTSHSTNLNNSLKLFEITLSLGNKNKIYQELIGQKSSSENFYER 969
 QY 891 RDKKVTLSKLFEGFKELQDEENFLRELSEKEMVDLQKMEATEVEKLFGRALPIRLK 950
 Db 970 LKSDDTYNESITTFVASKADDLNLSNDESKR-----KLE-EDINKL-----KKT 1014
 QY 951 MAVATGCTFCGHDRIFFSSDDGIGRLGTTKRINGTLLKILPPIQADLRTTGRSS 1010
 Db 1015 LQSFEDIYNNKYKLERLFPKRTVGYKKQKIKLT--LLK-----EQLESKINSIN 1064
 QY 1011 RP--LSAFRSGSGKGFIDVPLPSKNELEKELIAPLLKLVGLACAFIIVSADAVDAL 1067
 Db 1065 NEKHAVLQNFVFNK-----KKEAL-----IAEENLT 1092
 QY 1068 KTCACILKGR--IELACIANPACAANVACLOTNNRPDETECOIKC--GDLPFNSVY 1122
 Db 1093 ENTKILKHYKAVKYNGESSPLKLTSESIQTEDVYASLENFKYLSKLEGKLDKNLWL 1152
 QY 1123 DEFNFCANVR-----KKCVPRKSDLGEPAPDPSVLVONFISDENGKWTITSGLN 1173
 Db 1153 EKKKILSYISRGILHNLIALKEVINKNNTGNSPS-----VNTIDV-----N 1193
 QY 1174 PTFDAFQCLHEFTGDNKLVGNISWRIKTL--DSGEFFTRSAVQKRVODPQGVLYNH 1231

```

Db 1194 NALESY-----KFLPEGTD-----VATVSESGSTLLEQSO-----PKRA-----1231
QY 1232 DNEYLHXODWYLLSKINKPKEDYIFYYRGRNDA-WDYGAVYRSSLVPSIIP 1290
Db 1231 -SHVGAESTITTSQNVDEVDVYIVLIFGESEEDDGLQVYT---GRAVTSYV--1285
QY 1291 LEKAASIGRDFSTFIRIDNTGCPPEALVERIEKTEVEGERIIVKEVEIEEVEKEVEK 1350
Db 1285 -----DN-----ILSKIE---NEVEVYLKPLAGVYRSLKOLE-1316
QY 1351 VGRTEMLFORLAEGFNEIKODEENFVELSKSEKEVEFLD 1389
Db 1316 ---NNVMTENVNVKDLINSRKNREKNVLESDLIPYKD 1352

```

```

RESULT 14
R95268
ID R95268 standard; Protein; 993 AA.
AC R95268.
DT 01-AUG-1996 (first entry)
DE Pre-nisin modification gene nlsb product.
KW Nisin A; nlsb gene; antimicrobial; preservative; antibiotic;
OS Lactococcus lactis strain NIZO R5.
PN W09616180-A1.
PD 30-MAY-1996.
PF 20-NOV-1995; G02669.
PR 19-NOV-1994; GB-023404.
PA (BIOT-) BIOTECHNOLOGY & BIOLOGICAL SCI RES COUNC.
PI Dodd HM, Gasson MJ.
DR WPI; 96-268616/27.
DR N-PSDB; T29660.
PT Making cell which expresses nisin but does not contain natural nisa
PI gene - by providing cell with variant nisa gene, and genes for nisin
PT modification, secretion and immunity
PS Disclosure; Fig 7; 69pp; English.
CC The gene cluster nlsbAFCIPRK (see T29660 and T29661) of Lactococcus
CC lactis includes the nisa gene coding for pre-nisin A (R95267, see
CC also R95263) and the genes for nisin modification, secretion and
CC immunity. nlsb (R95268) and nisc (R95270) are believed to be
CC involved in reactions that modify pre-nisin; nist (R95269) is
CC similar to a transport ATPase and is involved in translocation of
CC nisin out of the cell; nisi (R95271) is involved in immunity to
CC nisin. Replacement of the natural, chromosomal copy of the nisa
CC gene with a variant nisa gene allows produ. of high levels of nisin
CC A variants in Lactococcus lactis hosts.
SO Sequence 993 AA;

```

```

Query Match 1.7%; Score 129; DB 1; Length 993;
Best Local Similarity 18.8%; Pred. No. 0.15;
Matches 125; Conservative 99; Mismatches 244; Indels 196; Gaps 28;
QY 297 LHYDDMYLLSSQENKPDY-----IFVYRGRNDAMDYGSVYTRSPTEPSTI 349
Db 171 IKTYNVQIISFECENDYQKEDICETVLCYQDEYRELSEYLSLT-----VNHILI 224
QY 350 PNLQKAKSVGRDF--NFFITNSGCPPEPLV---ERLEKTAEGEKKILLAEVEIEE 404
Db 225 SNLOKDLIS---DFSMNFIITKVEAIDEDKKYIIPLKRVQKIOEYSEIEGEGIEKE 281
QY 405 VEKE-----VEKVRDTEMLT---FORLEGEFE-----LQODEEN 436
Db 282 IYQESQILENDNYIQLIDLSSEINFQKQOQLEHLAEFLGNTKSVRTYLDYRKD 341
QY 437 FVELSKSEKEELINELQMAATEVEKIFGRALPIRKLRNALAHNSFLANHEITIKYVGSK 496
Db 342 FIKKYVDDEVQVTEL-----FDSFTGIGAPYNNH---PRNDFYSESTIYYSBE 391
QY 497 LPCHRRFSWGMEDYFSGIIVAKICSSRRIPRFRKSPITCGDLBSRGLQFSHRKHLNSP 556
Db 392 R-----EKYLSMYYEA-----YKNHNVINLDD 413

```

```

QY 557 ASINQNPKNKSGCKFPKRDVALMWEKMGQFATAIVATFIL-----SVASKDAVD 609
Db 414 LESHYQKM-----DLKSELOGLEFLNLAKYEYKDFILGDIYGNNGGASGRS 466
QY 610 AKTCTCLKEKRELELAKCISNPACAA-----VACIOTCNRPDETECOIKGDLFENS 665
Db 467 AL-----SPELTSYHRTIYDSVRENEKEIISCEIVF--LPEINR 505
QY 666 VDEFNECAVSRRKCVPRKSDVGDFFVPDPSV-----QKFMKDDSG---KWIITRGL 716
Db 506 HANVHTSIMARKVLPFFTSISNEVLLINVIYIGIDKEKFAARDISTQGEVLRKFTYSKY 565
QY 717 NPTDPAEDCQ-LHEFHTENKLVGNLSWRIRTPDGGFTF-----756
Db 566 NKTLEFSNELRPLIYISLDDK--FGNLPWELIYRDPYIRLIVDELIVISPAKAKTIMGROV 623
QY 756 ---SAVOKFVODPKYPIILY--MHNVEYLQD---DWITLSKYENSPEDIYV---Y 803
Db 624 NSKATIRLOSKETPKREFYIVNGDKVYLSCENPLDMEILSAIKSSKRKDFIELQY 683
QY 804 YKGRNDAMDYGSVLYTRSAVLPESIIPELOTAQKVRDNTFNIKIDNTGCPPEPLVE 863
Db 684 FEDENIINKGEKRV-----ADVVPFIRIRA--LQNGRAFIRKRVN-----VE 727
QY 864 RLEK 867
Db 728 RREK 731

```

```

RESULT 15
R77087
ID R77087 standard; Protein; 1582 AA.
AC R77087.
DT 31-JAN-1996 (first entry)
DE Rat sulphonylurea receptor.
KW Persistent hyperinsulinaemic hypoglycaemia of infancy;
KW sulphonylurea receptor; diabetes; COS.
OS Rattus sp.
PN W09528411-A1.
PD 26-OCT-1995.
PF 12-APR-1995; 004463.
PR 13-APR-1994; US-226972.
PR 15-MAR-1995; US-404531.
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
PA (TEXA ) UNIV TEXAS SYSTEM.
PI Bryan LA, Cote GT, Gager RF, Nelson DA, Thomas PM.
DR WPI; 95-373786/48.
DR N-PSDB; T01568.
PT Sulphonylurea receptors and corresp. nucleic acids - used in
PT detecting persistent hyperinsulinaemic hypoglycaemia of infancy
PS Disclosure; Page 99-104; 145pp; English.
CC Rat sulphonylurea receptor (SUR) DNA was ligated into eukaryotic
CC expression vectors containing the SV40 virus promoter and used
CC to transfect COS cells. Expression of the receptor was low
CC when SUR cDNA (T01561) was used, but high when the genomic SUR
CC sequence (T01568) was used; the product had the sequence given
CC in R77087.
SO Sequence 1582 AA;

```

```

Query Match 1.7%; Score 124; DB 1; Length 1582;
Best Local Similarity 18.8%; Pred. No. 0.79;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;
QY 868 KYEEGERITLIEVEIEBEVAKVDKVTLPFKLFGSKF--ELQDEENFLRELSEEND 925
Db 480 KLSQKQRTTL-----EYSNERLKQT---NEMLRKIKLKYAMENIFCSXVEXTRK 528
QY 926 VLDGKMEA--TEVEKIFGRALPIRKLMATAVATCFSPCHDRIRFSSDD-----GIGRL 978
Db 529 EXTSLRAFAVYTSISIMNTAIPAANVILFV-----GHVSFKESDPSVAFAVL 580

```

QY 979 GTRKRNIGFLL-----KILPIQ-----SADLRTTGGSSRPLSAFRSGFSKG 1023
Db 581 SLFHILVTPLEFLSSVSTVKALVSYOKLSEFLSSAEIREOCAPREPAPQGA----G 636
QY 1024 IFDVLPL-----PSKNELKELTAPLLKLVGLACAFILVPSADAVDAKTCACCLKG 1076
Db 637 KYQAVPLKAVNKRKRPAREEVVDLIGPLQ-----RLTPSIDG-DADNFCVOIIGG 684
QY 1077 C-----RIELAKCIANPACANAYACLQTCNNRPDETECOIKGCDLF 1117
Db 685 FFWTPDGIPTLSNITIRIPRGQTMIVGQVGGCKSSLATLG-----EMQKVSQAVF 738
QY 1118 ENSVYDEFNECAVSRKRCVPRKSDLGFEFPADP-----SVLVQNFNI-SDF 1162
Db 739 WNSLFPD-----SEGRRPOQPRAGDSGRFGCOEQRPCGYASQKPMILNATVEENITFESPF 793
QY 1163 NGKWY-----ITSGINPTFEDAFDQQLHEFHTEGDNKLVGNISWRIKT----- 1205
Db 794 NKQRYKAVIEACSLQPDIDIL-----PHGDOTQIGERGINLSTGGQRPDQCPEPST 845
QY 1205 -----LDSGF-----FTRSAVKRFVODPNQPGVLYNHNDEYLYHODDWTLLS 1247
Db 846 STPMIVFLDDPFSAIDVHLSDHLMQAGILELLRDKRTVYLVTHKLOYLPHA-DWIIAMK 904
QY 1248 KIENKPEDYLFVYRGRNDAMDGYGCAVYTRSSVLPNSIIPLEKAKSIGRDFSTFIR 1307
Db 905 DGTIQREGILKDFORSECOLFEHW-----KILMNRQDQELER-----ETVME 946
QY 1308 TDNTCGPEPALVERIEKTVEEGERIYKEVEEIEEVEKE 1347
Db 947 RK--APEPS--QGLPRAMSSRDGLLDEDEEEEAASE 981

Search completed: October 14, 1999, 03:56:10
Job time: 2693 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 1999, 02:42:34 : Search time 33.23 Seconds
(without alignments)
1702.452 Million cell updates/sec

Title: US-09-075-375-2

Perfect score: 7495
Sequence: 1 MALSHYVFLCKEALNLNYA.....MEASEVEKLGRALPIRKVR 1412

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database: PIR-60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2557	34.1	478	2	T03750
2	2430	32.4	462	2	T00708
3	155.5	2.1	886	2	H69378
4	155.5	2.1	2829	2	A42771
5	155	2.1	2469	2	H36812
6	146	1.9	1935	1	A37102
7	146	1.9	2748	2	S57976
8	144.5	1.9	839	2	S54174
9	144	1.9	2166	2	G70163
10	143.5	1.9	1558	2	B71603
11	142.5	1.9	1935	1	S06006
12	142.5	1.9	1819	2	A71928
13	141	1.9	1002	2	C70319
14	140.5	1.9	1302	2	JC6009
15	140	1.9	800	2	F64508
16	139.5	1.9	1025	2	S54044
17	137.5	1.8	1875	2	S38173
18	137	1.8	852	2	B43743
19	137	1.8	1676	2	E71410
20	137	1.8	1109	2	A40801
21	135	1.8	3660	1	S02041
22	134.5	1.8	1937	2	I38055
23	134	1.8	1639	2	S05603
24	134	1.8	978	2	A70387
25	133.5	1.8	1934	2	I48153
26	133	1.8	1979	2	C71622
27	132.5	1.8	1156	2	B70356
28	132	1.8	809	1	J00032
29	131	1.7	1156	1	E69444
30	130.5	1.7	1631	1	SAZOK1
31	130	1.7	1679	2	S48385
32	129.5	1.7	1805	2	A47297
33	129.5	1.7	1805	2	A64224
34	129	1.7	1175	2	A35815
35	129	1.7	1175	2	C35815
36	129	1.7	1822	2	S44849
37	129	1.7	1179	2	F71190
38	128	1.7	993	2	C31915
39	128	1.7	624	2	PC6003

ALIGNMENTS

Query Match	Score	DB 2:	Length	478:
Best Local Similarity	100.0%	Pred. No. 9	9e-151	
Matches 477:	Conservative	0:	Mismatches	0:
Indels	0:	Gaps	0:	
Query	474	MALAPSNLA-VTETIKYVSKLPGHRRFSWGMEDYFGSIIVAKICSSRRIRYRKSP	533	
Db	1	MALAPSNLA-VTETIKYVSKLPGHRRFSWGMEDYFGSIIVAKICSSRRIRYRKSP	60	
Query	534	RICCGDLSRCGLFSGKGNLSPASINONYPKNGSGCKPKDYALMWEKQFAKTAI	593	
Db	61	RICCGDLSRCGLFSGKGNLSPASINONYPKNGSGCKPKDYALMWEKQFAKTAI	120	
Query	594	VAFILISVAKADAVDALCTCLKEGRLLEAKCISNPACANAVACLOTCNNRPDETEC	653	
Db	121	VAFILISVAKADAVDALCTCLKEGRLLEAKCISNPACANAVACLOTCNNRPDETEC	180	
Query	654	QIRCGDLFENSVDENECVSRKCKVPKRSVGFPPDPDSVLYOKFDMDFSGKFFIT	713	
Db	181	QIRCGDLFENSVDENECVSRKCKVPKRSVGFPPDPDSVLYOKFDMDFSGKFFIT	240	
Query	714	RGMLPTDPAFVCHHEPTEENKLVNLSMRIRTPDGGFFTSVAVOKFVDPKTPGILYN	773	
Db	241	RGMLPTDPAFVCHHEPTEENKLVNLSMRIRTPDGGFFTSVAVOKFVDPKTPGILYN	300	
Query	774	HDNEVLYQDDWYTLSSKVENSPEDYIFVYKGRNDAMGYSVLYKRSVAVLPESITPE	833	
Db	301	HDNEVLYQDDWYTLSSKVENSPEDYIFVYKGRNDAMGYSVLYKRSVAVLPESITPE	360	
Query	834	LOTAQKXGRCTTFITKTDNTGCEPPLVERLKKVEEGERITIKVEEIEEVEKVRDK	893	
Db	361	LOTAQKXGRCTTFITKTDNTGCEPPLVERLKKVEEGERITIKVEEIEEVEKVRDK	420	
Query	894	EYTLSPKLFEGFKELORDENFRELSEKEMDYVDGLKMAEIVKELPRAPIPKKL	950	
Db	421	EYTLSPKLFEGFKELORDENFRELSEKEMDYVDGLKMAEIVKELPRAPIPKKL	477	

probable membrane
transport protein
probable membrane
glycoprotein A - m
probable major sur
nib protein - Lac

RESULT 2
T00708
Probable violaxanthin de-epoxidase precursor - Arabidopsis thaliana
N:Altenate names: protein F22013.3
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
C:Accession: T00708
R:Shim, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con-
eoligis, A.; Ecker, J.R.
Submitted to the EMBL Data Library, April 1998
A:Description: Genomic sequence for Arabidopsis thaliana BAC F22013.
A:Reference number: Z14200
A:Accession: T00708
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-462 <SH>
A:Cross-references: EMBL:AC003981; NID:g3063438; PID:g3063441
A:Map position: 1
A:Introns: 72/3; 128/2; 160/3; 292/2

Query Match 32.4%; Score 2430; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 6, 6e-143;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 951 MAVATHTCTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKTLIPPIQSDALRTTGRSS 1010
Db 1 MAVATHTCTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKTLIPPIQSDALRTTGRSS 60
Qy 1011 RPLAFRSGSGSKGFEDVLPSPKNEKELTAPLLKLVGLACAFLLVPSADAVDALKTC 1070
Db 61 RPLAFRSGSGSKGFEDVLPSPKNEKELTAPLLKLVGLACAFLLVPSADAVDALKTC 120
Qy 1071 ACLKGRICELACIANPACANAVACLOTNNRPDETCOICGDLFESSVDFNECAV 1130
Db 121 ACLKGRICELACIANPACANAVACLOTNNRPDETCOICGDLFESSVDFNECAV 180
Qy 1131 SRKCVPRKSDGEPAPDPSPVLYVONFNSDFNGKWTSGLNPTFPAFCQLEHFTGEG 1190
Db 181 SRKCVPRKSDGEPAPDPSPVLYVONFNSDFNGKWTSGLNPTFPAFCQLEHFTGEG 240
Qy 1191 DNKLVNISMRKTTLDGFTSRSAVQKEVODPNQGVLYNHNEVLAHODWYLLSKIE 1250
Db 241 DNKLVNISMRKTTLDGFTSRSAVQKEVODPNQGVLYNHNEVLAHODWYLLSKIE 300
Qy 1251 NKPEYIFYYRGRNDAMGCGAVYTRSSVLPNSIIPLEKAKASIGRDFSTFIRTDN 1310
Db 301 NKPEYIFYYRGRNDAMGCGAVYTRSSVLPNSIIPLEKAKASIGRDFSTFIRTDN 360
Qy 1311 TCGPEPALVERIEKTEVEGERIIVKEVEIEEVEKEVEKGRTEMLFORLAGFNEK 1370
Db 361 TCGPEPALVERIEKTEVEGERIIVKEVEIEEVEKEVEKGRTEMLFORLAGFNEK 420
Qy 1371 ODEENFVELSKEMEFLDEIKMESEVEKLFKALPIRKVR 1412
Db 421 ODEENFVELSKEMEFLDEIKMESEVEKLFKALPIRKVR 462

RESULT 3
H69378
Purine NTPase homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
C:Accession: H69378
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Nature 390, 364-370, 1997
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Atliach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Moese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343
A:Accession: H69378
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-886 <KLE>
A:Cross-references: GB:AE001032; GB:AE000782; NID:g2689355; PID:g2649562; TIGR:AF1032

Query Match 2.1%; Score 155.5; DB 2; Length 886;
Best Local Similarity 20.6%; Pred. No. 0.063;
Matches 137; Conservative 85; Mismatches 221; Indels 221; Gaps 27;

Qy 797 EDYIFVYKGRNDAMGCGSV--LYTRSAVLPDESTIPELQTAOKRGVDFNTFKTD-N 853
Db 152 EDY-----ENAKNKGAVIRKLEREKERKEFLFSOEQIKRQKEK-----KALE 197
Qy 854 TCGPEPALVERIEKTEVEGERIIVKEVEIEEVEKEVEKGRTEMLFORLAGFNEK 910
Db 198 RISEIKSIEISLREKLESEVRNLESRLKELEHKSLSLKRQESSVLQEV-----R 249
Qy 911 DEENFVELSKEMDV---LDGLKMEATEVERKLFGRALPIRKMAVATHTCTSPCHDRIR 967
Db 250 GLEKLELEKQKEVERIEDLEKKAKVEKEL-----KKAREY- 290
Qy 968 FFSSDDGIGRLGTRKINGTFLKTLIPPIQSA--DLRTTGRSSRPLSARSGSGIF 1025
Db 290 -----SLEKLSLSTINQALNDVKEKRGDLTREAAGIQAOQLKAAE 329
Qy 1026 DIVLPSPKNEKELTAPL-----LKLIVGLACAFLLVPSADAVDALKTC 1070
Db 330 D-----NSKLEITRKIELELELEKEREKSHRLE-----TLKPKDRMQGK- 373
Qy 1071 ACLKGRICELACIANPACANAVACLOTNNRPDETCOICGDLFESSVDFNECAV 1130
Db 373 -----AKLEKMLTPDKVE---KMDLLSKAKEE-KEITE 404
Qy 1131 SRKCVPRKSDGEPAPDPSPVLYVONFNSDFNGKWTSGLNPTFPAFCQLEHFTGEG 1190
Db 405 KLKLLAKSSSLKTRGAOLKRAVE-----LKSARTCPVCGRELDEHRRK 451
Qy 1191 -----DNKLVNISMRKTTLDGFTSRSAVQKEVODPNQGVLYN- 1231
Db 452 IMAEYTRMKRIAEFLAKADEIEKKLEKVEKALEKQETVLKROWDELKALENEL 511
Qy 1231 --HDEYLAHYDDWYLLSKTEKPEDYIFVYRGRNDAMGCGAVYTRSSVLPNSI 1288
Db 512 SSHDAE-----KLSAEESE---YRKVERLDLGRGQKILLSSA---SRI 550
Qy 1289 PELEKAKASIGRDFSTFIRTDNCGPEPALVERIEKTEVEGERIIVKEVEIEEVEKEV 1348
Db 551 KELKSLREI-----EAL-KVSESEKELHKKIREEGFEELELIEREV 593
Qy 1349 EKVRTETMLFORLAGFNF--ELKODENFVELSKEMEFLDEIKMESEVEKLFK 1405
Db 594 -----QSLRPYKMWLELKDAESRLSESLKRR-KLEDELSIAKLAEBAKGA 641
Qy 1406 LPIR 1409
Db 642 EEIR 645

RESULT 4
A42771
reticulocyte-binding protein 1 - Plasmodium vivax
C:Species: Plasmodium vivax
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: A42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:9231538
A:Accession: A42771
A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2829 <GAL>

A:Experimental source: Bellem strain, mercozites

A:Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBI:108115)

Query Match 2.1% Score 155.5; DB 2; Length 2829;

Best Local Similarity 17.4%; Pred. No. 0.33; Mismatches 573; Indels 529; Gaps 76;

Matches 283; Conservative 242; Mismatches 573; Indels 529; Gaps 76;

57 TSKTSSFSDDSHCKDKSQICSDTGFEEI-----QFDLKRGMTLLLEQWQFQOLA 110
 1110 TSKNNELKSVKEVEDKLMIVNEQNEDEYKVKKNPENKEKLELRGSMKLE----- 1161
 111 IVLVCTVYIPRDVADALKTACILKECRILELACIANPSCANVACLOTNNRPDETE 170
 1161 -----VIRKRVSEMTQLESTANTLK-----SNAKGKEHDELELN-----KTK 1199
 171 CQKCGDLFE--NSVDFNECAVSRKCVPRKSDVGEFVPVDRNA--VYQFNEM-KDF 224
 1200 GQMR--DIYEKKKIAELKEGTIVNELKDANEKANKE--PEPERNIGHVLERTIVEKD 1256
 225 SGKWTTSGLNPTFDACDLHFE--HMENDKLVGMLTRKITLDGGEFTRSVQTFVOD 282
 1257 AGK--VVEENMSLTKIKLELIQETSDSOSELVTTSTIHLNEMAKG----- 1301
 283 PDLGALYNHDFLHYODDMYILSSQIENKPPDYIFVYVYRGHN-DAMDGYGGSVITRS 341
 1301 -----YED-----VIRNEEDSIQLEKAKSLTLD-----EM 1328
 342 PTLPEIIPNLOKAK--SYGRDFN-----FITDMS----- 373
 1329 KTLVQVMMNLOSAILOGNAGISKELNKGVELLSTINSLILEYKANSESVAFSOL 1388
 373 GCEPRIVERLEKTAE--EGEKILIKEAV-----EIEEVEYK-----EVEKVD 415
 1389 ANGFTTAKGEKKNASALAEAKL--KEQIVKLDYSDIDDKVKIKLEGIKRELKMKES 1446
 416 EMILFQRLLEGFKELQDEENFVELSKEKEILN-----ELOME----- 456
 1447 ALTFWEE--SEKFKQWSSHMEVAKGKRRILEYLNNGDGGKANTDSQMEVEGVNYSKAE 1505
 456 -----ATEVEKLEFGRALPIFKLMALAPHSNLFANHEITIKY 492
 1506 HAHFIVEAOYDKTAFCEIYAVYTKMNLFNESI--MKEVAVKCEK-----KNDDEAKY 1559
 493 VGSKLPGHRRFSGWEDYFGSIVAKICSSRRIPRYFFKSPRICGLDSGLQLFSGHK 552
 1559 -SAKL-----KPYDGR-I-KARVSENERKISELEKAKV--EKRESSQINDVSTR 1603
 553 NLSBAHSINQVPRKNSGCKFPFDVALMWEKGOFAKTAIYALFILTSKADAVDALK 612
 1604 SLQIDRCROQLDSVLSTIGRYKQNLQYFDS--AKSKMSVLPISLELAEKSLDKVK 1659
 613 TCTCLKECRLELAKICSPACAAV--ACLOTNNRPDETEQ--TKGGLDFENSVYDEF 669
 1660 AA-----KESEYKMLETYQNMESRINVEEGSLNDIDKIKTIDINDLLKMKQYEEGLQKI 1715
 670 NECAVSRKCVPR-KSDVGEFVPVDPVSVYOKRDMKDESGKMFITRGNPTFDFADCOLH 728
 1716 KENDKRKSNFELVSGEINALDPSTISFI-KLKLKEYD-----MGLDKANKGVAMN 1766
 729 EHFHEENKLVGNLSMRI-----RTPDGGEFTRSVAVOK 760
 1767 EIHGEFTSYNLIETHLSNMTDVSVPTEKAQSLRELAKEEHEHLRREREELAILLNDIK 1826
 761 FVQ-----DKPIPGILYNHDENYLLYODDMYILS----- 790
 1827 VESLKLKEMMKVSAEYEGKRRDHTVSQLYODMKTYDELKTLNDISEGSSVLNNVYS 1886
 790 ---SKYENSPD-DYIFVYVYKGRNDAMDGYGGSVITRSVAVLPESIIPELOTAAOKVGRDN 846
 1887 IVKVKESHADY-----KRDANSMYESAVTLANYFLSDEKAKISS-----GMEFN 1931

847 TPIKTD-----NTCGPEPPLVERLE-----KVEEGER-----TIKEV 880
 1932 AEMKSNKXTDLIELIFSVISNSNELKKIIEODSNDVIOKEHESQGLKADTDIYVAKLK 1991
 881 EEIEEVEYKCKEVTILFSKFEKFELODE-----ENFLRELSEEDVLDGLKMEA 934
 1992 NEFNEKELEAKNNEEVSVKREALKRLSOVEGIRCHENFRLDTE----- 2041
 935 TEVEKLEGRALPIKLMATAVHTFTSPCHDRIRFSSPDGIGRGITRRKINGTFLLKIL 994
 2041 -ELENTL-----KKVVTY-----RKKSERESGLOEM--ENEMN-TYSNSI- 2078
 995 PPIOSADIRTTGGRSSRPISAFRSGFSKGFIDVPLPSKNE--LKLTLPLLLKLVG 1050
 2078 --TQLEGIVASAGSK-----DIEKLENSNEMRNISERISTDSK----- 2118
 1051 LACAFILVPSADAVDALKTACILKCGRIELAKCIANPACANVACLOTNNRPDETEQ 1110
 2118 -----VIENMSI-DELYK--LGKNCQAHWISLSLYTANKTISKLMINKENKTE-- 2167
 1111 IKCGDLFE--NSVDFNECAVSRKCVPRKSDIG--EPAPDPVLYONFNISDFNGKY 1167
 2167 -KCVDTIKDNSSSTDGYVELT-----KGFYGSKLTFSSASEIYQADTVSYNFAKH 2216
 1168 ITSGLNPTFDACDLHFEHTEGDKLVGNISMRKITLDSGFEFTRSVAVOKFVQDPNOGV 1227
 2217 EKESLNATRD--IKELLYLFHONSISIV-----EGGVOML-----A 2252
 1228 LYNHDFLHYODDMYILSSQIENKPPDY-----IFVYVYRGND----- 1268
 2253 LVDKLINEKRENDLYRNISETKLKOMEHSDVFKPIELHAKMNETNNKSLLEKELKL 2312
 1268 -----WDGYGAVVYTRSSV--LPN--SLIPLLEKAASTORD----- 1302
 2313 SYNDHMSMEA-MIKNGKLTYPESVONINNTYVIEAEVTLIEDIDYDNTQYVEEK 2372
 1302 --FSTIRIDNCGEPDLVERIEKTEVEGE--RIIVEVEIEEVEEVEYK--GR 1353
 2373 KQPSILIDRTN-----ALMDIEIFKKNYNNLMENVEITLHVVDYIYKIKTKLVQAK 2426
 1354 TE-----MTLFGQ--LAEGFNEKODEENFVRLSKEEM-----EFLDE 1390
 2427 TEVEQILENIKQMDMLQNIPLKQVSIIEYFENVKKKESILNDLYBOERLLKIGEHLD 2486
 1391 TKMEASE 1397
 2487 IKRWATE 2493

RESULT 5
 H36812
 Hypothesized protein ORF64 - saimiri herpesvirus 1 (strain 11)

C:Species: saimiri herpesvirus 1
 A:Note: host Saimiri sciureus (common squirrel monkey)

C:Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text-change 09-Sep-1997

C:Accession: H36812
 R:Albrecht, J

Submitted to the EMBL data library, January 1992
 A:Description: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A36806
 A:Accession: H36812

A:Molecule type: DNA
 A:Residues: 1-2469 <A,B>

A:Cross-references: G3:564346; NID:960320; PID:960385
 R:Albrecht, J.C.; Nichols, J.; Biller, D.; Cameron, K.R.; Blesinger, B.; Newman, C.;

J. Virol. 66, 5047-5058, 1992
 A:Title: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A37309; M01D:9233568
 A:Contents: annotation; protein-coding frames

A:Note: neither protein nor nucleotide sequence is given

C:Genetics:
 A:Gene: 64

Query Match 2.1% Score 155; DB 2: Length 2469;
 Best Local Similarity 17.3%; Pred. No. 0.29;
 Matches 230; Conservative 208; Mismatches 497; Indels 396; Gaps 58;

QY 271 FTSSAVOTVYVODDPLFGALYNHNDNEFLHYODWYTLSSQINPKRDDDIYFYR-----324
 DB 178 FTSNVDNDAIEYLSPPNVOYT--GSFLYFVKREYIGHSH-----YIMNHYRINYEKU 228
 QY 324 -GENDAMDGYSVYI-----TRSPITLPESTIENLOAKASVGRDENFI 367
 DB 229 HGNNDILTSOEBLIIIFISPPNTPKPTSTOKPKPTPKPTKATKAPKPTPKPTKPTKPT 288
 QY 368 TIDNSGPEPLVERLEKTAEGEKLLIKKAV-----EIEEVEKKEVEVROT--EMTL 419
 DB 289 PYDKS--KKP--KIPKTSKSKKVLTKDIALPQKITEHIRELLPPIETVEDNTL 343
 QY 420 FOR-----LLEFKELQ-----ODENFVRELSKEKEILINLOMEATEVEKL 462
 DB 344 FHHPERTTPGIDSLGINSITTKREDLDDDDVNTSKLEDEDDWIDD-----393
 QY 463 FGRALPIKRLMALPHSNFLANHEITIKYVGSKLGHKREKSGMEDYGSIVA-----518
 DB 393 -----IPIEVLDTETHS--DOETI-YMIGDE--NIHDMSTDDDDIDDLIDISFIQID 441
 QY 518 -KICSSRRIPRYFRKSRICCGLDSSGLQLEFSHGK-----HNLSPASINONVPG 567
 DB 442 NLITSLDNIPRN--NTEPRI--IDKTSNOPIKKGKALHSIDRLKNIVLEHGLINSSIS 497
 QY 568 NSGCKFPKDVAMVMEKQOFAKTAIYAFILTSVASKAVALATCICLKECKELAK 627
 DB 498 ISKSKSLQEVYL--WGE-----KLSIPTR-DLKITILLITELITELIAELATLK 542
 QY 628 CISNPAACANACIQOTNNNPDETECOIQCGDLFENSVDDEFNECAVSRKCKVPRKSDVG 687
 DB 543 -LIN-----DIFRNINVTIKKCMMLK-----SSVD 569
 QY 688 DEF--VPPSVLYOKFDM-----KDFS-----GKPFITGLNPTF 720
 DB 570 SYKHLASLNNITLIKQIDTELTELSTVTTSELGDFSVCTKRESEITMAIAIKLK 629
 QY 721 DARDQOLHEHTENKLVGNLSMRJRTPDGFFTRSAVOKF-----VODPKYPG 769
 DB 630 EKSTKQELHEENFOSVL-----IAMEIQPIPLPRVIELOPSKAAQ 675
 QY 770 ILYNHDNEYLLYODDWYIILSKVENSPEXYIFVYKGNND--AWDYGGSVLYTRSAVL 826
 DB 676 QL--HEKSKLVGCK--LTIDANNVLLDLHTMKODTIDISPAVD-----FTVLKN 722
 QY 827 PESIIELOTAOKVGRD--FNFIKIDNTGCEPPLVERLEKKEVEGERIT--IKEVEE 882
 DB 723 IOSTLOLOTCVTDINIKKFISSNTVOQLSYIGWEVAFELSGQNFKADPVLPVLTID 782
 QY 883 IEEVEKVDKREVLFLSKLFEKFELODENFIRELSKEDMDVLDGLKEATEVEKLEFG 942
 DB 783 IKKEIQOYITKO--KNEETLSILADVOYLLLENAKOSDTISPILOHYITKAGTLVG 837
 QY 943 RALPIRLKMAVAVHCTSPCHDRIRFFSSDGLIGLIT--RKINSGFLIKITPIQ 998
 DB 838 E-----RENQFESKINVTOKLSTSEFIKITLIDSTILENVOLOIQ 878
 QY 999 S-ADLRITGGRSSPLAFSGFSK--IFDIVPLPSKNEKELTAPLL-----1046
 DB 879 EISDLQSNQYIHQSETIKQAFEDKSNITINNLILOLINOQKYTYVQPLVAVKREFLSA 938
 QY 1046 -----KLGVLAACAFILVPSADVADALTKACCLLGGRIELAKCIANACANAC 1096
 DB 939 KFRSNTICEIITSLVGLSLKSTIVETAKALASTITLLEKL-----984
 QY 1097 LOTONNRPDETE--QICCGDLFENSVDDEFNECAVSRKCKVPRKSDGLGFPAP 1149
 DB 984 --TAVDRPLKRELYNVIKRLQOKLTKLLHQEFDMW-----KNEVSPF-VPT 1027

QY 1150 PSVLYONF--NISDFNGKYYITSGLNPFDFAEQOLHEFHEGDN-KLVGNISMRJKTLD 1206
 DB 1028 PSRDVATFTIOMASMKAKOYAKKALKDQIOAMEIDVDPESVIEDNIKANOKKMO-----1083
 QY 1207 SGFFTSAVOKFVOPDNOGVLYNNHNDNEFLHYODWYTLSSQINPKRDDDIYFYR-----1083
 DB 1083 -----KIQSNFQDLNLSIL-----PDDWLSL-AKEYTRKSTLFTVIGPILL 1124
 QY 1262 RGNNDAMDGYSVYI-----TRSPITLPESTIENLOAKASVGRDENFI 367
 DB 1125 KVEEVEKKEVEVROT--EMTL 419
 QY 1297 SIGRDSSTIRFNDT--CGEPALVERIE-----KIVE-----EGERITVEK 1337
 DB 1185 NIGHELSTLSQALNKSTIPPEAVGTSLQHAANFSCMFETLEATWHDQVDRTRKIDEYI 1244
 QY 1338 EIEEVEK-----EVEKVR--TEMILFORLAGFVNLKODEENFRELKEMEFL 1388
 DB 1245 EDLRNDTRKHIVAPQIOSPNRFLSPEDIQEINSLPKLFRDSLIENESRLASQKNEFOML 1304
 QY 1389 DEIKMEASEVE 1399
 DB 1305 -ENYVKAHEIQ 1314

RESULT 6
 A37102
 myosin beta heavy chain, cardiac and skeletal muscle - human
 N:Contains: myosin ATPase (EC 3.6.1.32)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence, revision 31-Dec-1993 #text, change 20-Mar-1998
 R:Accession: A37102; S12733; A94224; B28908; A24997; A27858; I54254; S12458; S09331;
 R:Juaenike, T.; Diederich, K.W.; Haas, W.; Schleich, J.; Dichter, F.; Pfordt, M.; Bac
 Genomics 8, 194-206, 1990
 A:Title: The complete sequence of the human beta-myosin heavy chain gene and a compar
 A:Reference number: A37102; MUID:91065634
 A:Accession: A37102
 A:Molecule type: DNA
 A:Residues: 11935 <JAN>
 A:Cross-references: GB:M57965; GB:M30603; NID:q179507; PID:q179508; GB:M30604; GB:M30
 R:Liew, C.C.; Sole, M.J.; Yamauchi-Takahara, K.; Kellam, B.; Anderson, D.H.; Ian, L.;
 Nucleic Acids Res. 18, 3647-3651, 1990
 A:Title: Complete sequence and organization of the human cardiac beta-myosin heavy ch
 A:Reference number: S12733; MUID:90301496
 A:Accession: S12733
 A:Molecule type: DNA
 A:Residues: 1-106, 'E', 108-671, 'LYH', 675-857, 'A', 859-941, 'NV', 944-1123, 'A', 1125-1158,
 A:Cross-references: EMBL:X52889; NID:929726; PID:929727
 R:Yamauchi-Takahara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
 Proc. Natl. Acad. Sci. U.S.A. 86, 3504-3508, 1989
 A:Title: Characterization of human cardiac myosin heavy chain genes.
 A:Reference number: A94224; MUID:89264452
 A:Accession: A94224
 A:Molecule type: DNA
 A:Residues: 1-87, 'Q', 89-106, 'E', 108-177, 1325-1702, 'DR', 1705-1786, 1788-1803, 'E', 1804-1
 R:Yamauchi-Takahara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
 Proc. Natl. Acad. Sci. U.S.A. 86, 7416-7417, 1989
 A:Reference number: A94226
 A:Contents: annotation; erratum
 R:Kurabayashi, M.; Tsuchimochi, H.; Komuro, I.; Takaku, F.; Yazaki, Y.
 J. Clin. Invest. 82, 524-531, 1988
 A:Title: Molecular cloning and characterization of human cardiac alpha- and beta-form
 human atrium.
 A:Reference number: A92770; MUID:88299163
 A:Accession: B28908
 A:Molecule type: mRNA
 A:Residues: 1412-1518, 'R', 1520-1574, 'NV', 1577-1935 <KUR>
 A:Cross-references: GB:M21665
 A:Note: The authors translated the codon AGC for residue 108 as Arg
 R:Uchiter, P.; Umeda, P.K.; Levin, J.E.; Vosberg, H.P.
 Eur. J. Biochem. 160, 419-426, 1986
 A:Title: Partial characterization of the human beta-myosin heavy-chain gene which is

Query Match: 16.9%; Pred No. 0.74; Mismatches 529; Indels 434; Gaps 51.

Best Local Similarity: 207; Mismatches 529; Indels 434; Gaps 51.

Matches 238; Conservative 207; Mismatches 529; Indels 434; Gaps 51.

203 DVEFFVPPRRNAVQ-----NNKKDFSGKMYTTSGLNPTFDPADQOLHER 248

461 DIAGFELFEPNSFEOCLINFTEKLOEFPNNHNFVLEDEYKKEGIEWTFIDFQMDLOAC 293

249 HMEUDKLVG-----NLTWIKTLDGFFTRSAVQFVODPDLPGALYNHD 348

521 IDLIEKPMGLMSLLEECPMPATMFFKALFD-----NHL 557

294 NEFLHYODDWYLLSSQLEKRP-DYTFYVGRND-----AMDYGSVITYTSPILPESI 348

558 GKSANFOQ-----PRNIGKPEAHSLIHVAGIDVYNIIGW-----LOKKNDPLNETV 605

349 IPNLOKKA-KSVGRDENFTTTDNGSGPEPLVLEKTEAEGEKLLIKEAVEIEEYVK 407

606 VGLYXKSLIKLSTLFANY-----AGAAP-TEKGKKAKAKSSFGQVSALH-BENLKK 657

408 EYEAKPDRDENTLFLQRLLEGFKELDQDENFVELSKKE-----LLNLOEAT- 458

658 LMTNLST-----HPHFVCIILPNETKSGYMDNPLYMHQLRONGVL 699

458 EYKLEGRALP-----IKLIMLAPHSNLANHETIKYVGSGLPGHKRPSNG 506

700 EGIRICRKFPPRILYGDFFRORYILPALIPGQFIDSRKGAELKSLSDIDINQIKFG 759

507 WEDEGSIYVAKICSRRIIPRYFKRSPICCGDSGIDOLFSGHNHSPAHS---INON 563

760 HTKYFFAGLIGLDEMDERLSRIITRI-QAOSRVLARMEYKLLERDSLSLVIOWN 817

564 VPKNSGCKPRKYVALAMWEKMGQFAKAJAIYVILTSVASKADAVDAKTCILKLEC- 622

818 IR-----AMGYKYN-PMKKLYFIKIKLSAREKEMASKEEFTRIKLE 864

622 -----RLEAKCISNPACANVACIQTCNNRPDEIOIKCGJLFNSV-----VDFENE 671

865 KSEARRKLELEKKNVSLLOEKNDLOLOVQAEODNLADAEHRDQIKKIKTOLEAKYENNE 924

672 -----CAVSRK--KCVPRKSDVGDFVPDPVSYLVKQFDMKDKSGKWFITRGL 716

925 RLEDEEMANBELTKKRKRLEDECSSELRDIDLELTLAKYERK----- 969

717 NPTFADFDCQHLHERHTEKNLYGNLSMRIRPTDGGFTTRSAVQKRVODPKYPCGLYNHDN 776

969 -----HATEENK-VKNLITEMAGD-----ETIAKLTKEKHALQEAHQAL 1007

777 EYLLYODDWYLLSKVENSPEDEYLYVYKGRNDAMGYGSGVLYRSA----- 825

1008 DDLOAEBDKYNLTLKAKYKLEOV-----DDEGSLSQEKKYRMDOLEAKRKL 1056

825 ---VLPESITPELOCTAAQV-----RDF-----NTFIKTNDTCPE-----PLVE 863

1057 GDLKLTQESIMDLENDKQOLDERLKKKPFELNALNAIEDEQALGSOLOKIKELQARIE 1116

864 RLEKVEEGERT-----IKVEEIEEVEVK----- 890

1117 ELEBEL-ESERTARAKVEKLRSDLSBELSELLELAEAGATSVQLENNKKREAEFOKMR 1175

890 -----VROKEVTLFSKLEFGFKELQDDENFILRELSKEMVLOGLKME 933

1176 RLEBALTQHEATAALRRKHADSVALEQOINDQIRKQKLEKSEKPELDED-----V 1231

934 ATEVEKLFGRALPIRKLMVAATCPSPCHDIRFFSSDDIGRLGITTKRINGTFLKI 993

1232 TSNMEDILIKAKANLEKM-----CRF-LEDOMNHRHS-----KAETQSVNDL----- 1274

994 LPIQSGADKNTYGGSSKPL-----SARRSGESGIDIVPLPSKMLKELTAPLLKLVGY 1051

1274 --TSORAKLOTENGELSRODEKREALISOLTRG-----KLITYOOLEDKLROLEEBYAK 1321

Thu Oct 14 07:57:25 1999

us-09-075-375-2.rpr

Page 6

QY 1051 LACAFIYPSADAVDAKTCACLLKGRICELACIANPACAAVACLOTQNNRPDETCQ 1110
Db 1327 NALHAHLOSARHDCDLLEQYEEETEKALLOVLSK--ANSEVAQMT----- 1374
QY 1111 IKCDDLEFNSVDEPNCAVSRKCYPRKSDGEEFPAPDPSVLQNFNISDFNKKWITS 1170
Db 1374 ----KYEIAIQRTETLEAKKRLAQRLOEAE-----AVE 1405
QY 1171 GLNPTDAFDCQLEHEFTGDNKLVGNISWRIKTLDSEFTFSAVQVQFVODPNQGVLYN 1230
Db 1406 AAVAKSSSEKTKHRLONLEIDLVND-----VERSAAMAAADKKO---RN 1448
QY 1231 HONEYLHYODDWYILSSKIENKPEYIFVYGRNDAMDGYGAVVYTRSSVLPNSLIPE 1290
Db 1449 FDKILAEWKQYEESESESSQKE-----ARSLTEFLKKNAYEES 1491
QY 1291 LE-----KAASIGRDTSTFRDNTGCPALVERIEKTEVEGERIYKEVEIEEBV 1344
Db 1492 LEHLETFRRNKNOLEELISDTLEQIGSSCKTITHELEVKRQLEAKEMELQSLAEAEASL 1551
QY 1345 EKEVEKVGRTETLFO-----RLAGEFNEKODEENFVREL----- 1382
Db 1552 EHEEKKILRAQLEFNQIKAEIKELAEKDEMOAKRNLRAVDSLQTSLAETRSMEA 1611
QY 1382 ---KEENE-FLDEIKMASEVEKLFPGA 1405
Db 1612 LRVAKKMEGDLNEMEIOLSHANMAEA 1639
RESULT 7
S57976
Nuclear migration protein NDM1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YD8358.06; protein YDR150W
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1996 #sequence revision 01-Mar-1996 #text-change 06-Feb-1998
R:Murphy, L.; Richards, C.; Harris, D.
submitted to the EMBL Data Library, July 1995
A:Reference number: S57971
A:Accession: S57976
A:Molecule type: DNA
A:Residues: 1-2748 <MUR>
A:Cross-references: EMBL:250046; NID:9899393; PID:9899399; MIPS:YDR150W
R:Kornman, J.; Schaeff-Gerstenschlaeger, I.; Zimmermann, F.K.; Perecko, D.; Kuentzel, H.
Mol. Gen. Genet. 230, 277-287, 1991
A:Title: Nuclear migration in Saccharomyces cerevisiae is controlled by the highly repet
A:Accession: S19052
A:Reference number: S19052; MID:92079907
A:Molecule type: DNA
A:Residues: 1-1569, 'V', 1571-1821, 'K', 1823-1959, 'RH', 1963-1970, 'RN', 1973-2048, 'N', 2050-2
A:Cross-references: EMBL:X61236; NID:94071; PID:94072
A:Genetic: SGD:NDM1
A:Cross-references: SGD:S0002557; MIPS:YDR150W
A:Map position: 4R
Query Match 1.9% Score 146; DB 2; Length 2748;
Best Local Similarity 17.9%; Pred. No. 1.2;
Matches 283; Conservative 232; Mismatches 581; Indels 486; Gaps 72;
QY 62 SFSFSSSSCKKQSCQSDISFEIQRFDLKGWTLLEKQMF-----IQLAIVLC 115
Db 748 SAYEEDLVKCKENPDV-----EFLKRSKAKLHTIVASSEYSELORKYSELEKEVEPS 800
QY 116 TFIYPRYDAVDALKTACCLKEGRLELACIANPSC-----AAVACLOTQNNRPDET 169
Db 801 LAYIVHAHAKTIDH-----HLSDSAYEELVKCKENPDMEFLKRSKAKLHTIVASSEYSEL 856
QY 170 ECGQ-----KCGD--LFNSVVDQENECAVSKKCVKPSVGEFPVDPDRNAV 215
Db 857 EKKLEPSTAYIVLVEHAKATDHHLLSDSAYEDLVK-----KENSQV-EF----- 900

QY 216 VQNFNMKDFSGKWTITSGNLPTDFADFCOLHEFHENDKLVGNLTWRKITLDGFFTRSA 275
Db 900 ----LKEKSAKLGHTIVSNEAYSELEKLEQ-----PSLAYIVHAHAKATDHHLLSDSA 948
QY 276 VQTFVODPDLPGALYHNDNEFLHYO-----DDWYILSSQIENKPPDYTFVYGR 324
Db 949 YE-----DLVKCKENPDMEFLKRSKAKLHTIVSNEAYSELEKLEQPSLEYIVHAHA 1002
QY 325 RND--AMDGYGVS-----IYRSPTEPSIIPN-----LOKAKSGRDF----- 364
Db 1003 TNHLLSDSAYEDLVKCKENPDMEFLKRSKAKLHTIVSNEAYSELEKLEQPSLEYIV 1062
QY 364 ----NNFTTDS-----CGPEPPIVERLEKTAEGEKLKAV-EIEEVEKE----- 409
Db 1063 HAKATNHLLSDSAYEELVKCKENPDMEFLKRSKAKLHTIVSNEAYSELEKLEQPSLE 1122
QY 409 --VEKVRDTEMTLQRLLEGKELQODEENFVRELSEKEKELINE--LOMEA-TVEKLF 463
Db 1123 YLVEHAKATNHLLSD--SAYEELVKCKENPDMEFLKRSKAKLHTIVSNEAYSELEK-- 1179
QY 464 GRALPIKRLMALAPHSNPLANHEITIKYVGSKLPGHRSFGWEDYGSIVAK----- 519
Db 1179 ----KLBO--PSLAYIVH-----AKATDHHLS--DSAYEDLVKCKENPDV 1217
QY 519 ----ICSSRRIPRYFRKSPRICCGDLSRGL-OLFSHGKHNLSPAHSINQW 564
Db 1218 EFLKRSKAKLGHTIVSNEAYSELEK-----LEOPSLAYIVHAH--ATDHHLLSDSA 1268
QY 565 PKGSGCKFPKQVDAVMWKEKMGOFKTAIVAFILVASKRD-----AVDAKTCQTC 617
Db 1269 YEDLVKCKENPDMEFLK--EKSALGHTIVSNEAYSELEKLEQPSLEYIVHAHAKATNHLL 1327
QY 618 LKRCRL-FAKCSNPAC-----AAVACLOTQNNRPDETCQIGDIFENSVDEN 670
Db 1328 LSDSAYEDLVKCKENPDMEFLKRSKAKLHTIVSNEAYSELEKLEQPSL----- 1378
QY 671 ECAYSRKCKVPRK-SDVGF-----PVPPSYLVOKFMDKDPGSGWFTTRNLNPTFAF 723
Db 1378 EYLVHAHAKATNHLLSD--SAYEELVKCKENPDMEFLKRSKAKLHTIVSNEAYSELEK-- 1435
QY 724 DCQ-----LHEFH-----TEENKLVGNISWRITPDGFFTRSAVOKFOD--PKYGLILYN 773
Db 1436 DYELLRSKLGKYLHIDTITNEIYSNFN-----SPLTKFIEEAKSKGRRLI 1482
QY 774 HONEYL-----LYODDWYILSK-----VENSPP-----D 798
Db 1483 EPNEYLDLNIATTPSKKEIDNECKQIGCYALDSKEYERLKNLSLENSKKEFEENALLD 1542
QY 799 YIEV--YKGRNDAMDGYGGSVLTRSAVLPESTIIPLOTAAOKVRDPTFTKTQNTC 855
Db 1543 LVLVDKTEYQAMD-----NASKKSLIPSTKAL-----DEVIM----- 1577
QY 856 GPEPPIVERLEKKEVEGEFTTIKEVEIEEVEKVRKDEVTLSKLFEGKELEQD----- 912
Db 1577 -PAPQLASAKESLSQK-----RTLSDIENELKAL-----GVAIKREKPN 1616
QY 912 -BENFIRELSKEEM-----DVLDDGLKMEATEVEKLFQ-----PALPIR 948
Db 1617 LKRPYDNASKNDVNLNCSKPSLVPLSTEEYDNNKREHKLINILDDPSIDPLKCKEKY 1676
QY 949 KLMAVATH-----CTSPCHDIRFSSDDIGRLGTRKRINGTFLKLTPPIOSAD 1001
Db 1677 QWLISKIDYEKQEALENPGYEFILEKASALGY-----ELVSEVELDRKKQKQIMSPD 1729
QY 1002 LRTTGGRSSRLSAFRSGFSKGFIDVPLPSKNELEKELTAPLLKLVAVIACAFLIYPSA 1061
Db 1730 IDYMOEKAAAN-----EMVLL--KNEKE-----ALOKKIEPPLFLFLIKAA 1770
QY 1062 ----DAVDALKTACCLKGRILELACIANPACAAVACLOTQNNRPDETCQIKG 1114
Db 1771 GNMKILVDQIEYDET-----IRKC-----NHFTRMELBESCH 1802

QY 1115 DLFENSVDNEFNECAVSRKKVPRK--SDLGEPAPDPSVLVONFNISDENGKWTISGLN 1173
 Db 1803 HL--NLVLLDQNEYSTLEBLENRNVEDINTLSKNTAIPTIYDOLGKYE-----N 1855
 QY 1174 PTADADCCQHEHTEGDKLVGNISMRIKTLDSGFTSRVAKQFVODPPQGVLY----- 1230
 Db 1856 PNFEDYKDSLNK-----MDYVAISRODYELMVAKKYKRPOLDYIKIS 1896
 QY 1230 -----NHDNEYLHODDWTYILSSKIEKKPEDEYIFVYVYRGNDAMDG 1270
 Db 1897 SEKIDHIVPLSEYNLMWNTYRNPSTLYKEKAVLNHLLIKEDYKNIL-----AVSE 1950
 QY 1271 YGGAIVYTRSSVLPNSIIPLEKAKSIGRDFSTFIIDNTCGPPALVEKTEVEGE 1330
 Db 1951 HPTVILHSEKASLLNKLVDKD-----DFATMSRSIE--KPTIDFLSTKALSMG- 1998
 QY 1331 RIIVKEV---EEIEEVEKEVKEVGRTEMFLFOLAGFENELKODENFVRELKPEEM 1385
 Db 1998 KILVNESTHKKRNEKLEPSEFLTKAKBOGLIITSEKYSRLRQIDRPSLDYLKEKA 2057
 QY 1386 EPLDEIKMEASEVEKLFEGKALP 1407
 Db 2058 AIFDSIIVENIFYOQLVNTTSP 2079
 RESULT 8
 S54174
 topoisomerase I - Plasmodium falciparum
 C:Species: Plasmodium falciparum
 C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
 C:Accession: S54174
 R:Tosh, K.T.; Kilbey, B.J.K.
 Submitted to the EMBL Data Library, January 1995
 A:Description: Isolation and characterization of the topoisomerase I gene from plasmodium
 A:Reference number: S54174
 A:Accession: S54174
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1839 <TOS>
 A:Cross-references: EMBL:X83758; NID:g790481; PID:g790482

Query Match 1.98; Score 144.5; DB 2; Length 839;
 Best Local Similarity 17.38; Pred. No. 0.28;
 Matches 191; Conservative 176; Mismatches 390; Indels 347; Gaps 46;

QY 352 LOKAKSVGRDNNFITDNSCGPEPLVERLEKTAEGEKLKAEVLE--EEVEKEV 409
 Db 24 INKIKONLG---NN--KSCNSRSSKESIKKOKNSSELGIKTKTKSLGKKEEKKKQI 78
 QY 410 EKVRDEMTLFLRLBEGRELOODENFVRELSEKEKELNLOMEATEVEKLFGRALPI 469
 Db 79 SKRKSSELKKNKKEKKRYEKKSRIV---KDETLLTVIKETQNNK-----PK 128
 QY 470 RKLRLMLAPSNFLANHETIKYVGSKLPGHKRFSWGEDYFGSIYVAKICSSRRIPRYE 529
 Db 129 KLLKLS---EENF---EPINRW--WEKIDQDQIDQNMV----- 159
 QY 530 RKSPRICGLDGRGLQFSGKRNINSPAHISNONPKNSGCKFPKDYALVWMEKQOFA 589
 Db 159 -----LEHRLG-LF-----SPPY-VQHHV-----IFYKSIKIEL-----NA 188
 QY 590 KTAIVAFILSVASKADAVDALKTCTCLLEKRLAKLISNPACAAVACIOTCNRNPD 649
 Db 189 KSEELATWCS-----AIGSDYCTKEKFIINEFKTFIN-----SIENNIITKO 231
 QY 650 ETECOIKGDLFENSVDNEFNECAVSRKKVPRKSDVGDPEVPDPSVLQKTFDM----- 704
 Db 232 ENETKLLKGDLSINEKFI-----FMPIDKHLLKREELKNTKTEE 271
 QY 704 KDESGKWFITRGINLTFPADCQQLHEFTEENKLVGLMSRIKTPDOG-----FTRSAY 758
 Db 272 KEERKMRKEKELPYTAYALVWIRKISSNKAEPPLGFRGGEHPKGLKKRIFFEDVY 331

QY 759 OKFVODPKYPCGLYN-----HDNE--YLLYODDWTYILSSKIEKSPEDYIFVYVYK 806
 Db 332 INTSKADVPRLYDNNCGHMGDIYDNKVTWLVAYKD-----SINDOIKYTFIS 381
 QY 807 RNDAMGCGSGSYLYTRSAVLPESIIPELOTAOKVGRDENTFIK-----DNTGPEPEPLV 862
 Db 382 AOSKEKGYKDI-KYENAR-----KLKSCVHKIREDYKRNKKNKNTIDQJGTAVYLI 433
 QY 863 ERLEKVEEGERTIIKEVEIEEVEKVDKREVLTFSLKEFGFELORDEENFRELSEKE 922
 Db 434 DELALRV--GGEV-----DIDEADTV-----GCCSLRVEHISFAHDIPFK 472
 QY 923 EMDVLDGKMEATEVEKLFGRALPIRKLMAVATGCTS---PCHRIFFESSDGIIRLG 979
 Db 473 SYD---SKEKTDKXKNNKIPLTNLESISSEDCIITLDFLGKSIIRFNT----- 521
 QY 980 ITRKRINGTFLLKTLPLQASDLRTTCGRSSRPLSAFRSGFGKGLFDIYPLPSKNE-LKE 1038
 Db 521 ---VKIDQAVINII-----IFCKNNRDEGVFDQITCSKLNELYKE 559
 QY 1039 LIAPIILKLVGLACAFIIPSDADVDAL--KTCACILKGRILELAKCIANPACAAVAC 1096
 Db 560 IMPTLSAKVFTYNNASITLDDQLKRIKRYGKITYSLSYG-ETELHK-----SK 607
 QY 1097 LOTNNRPDETECOIKGDLFENSVDNEFNECAVSRK---KCVPRKSDIGEPAPDPSY 1152
 Db 608 KRKSSHLTSDTNILSDASSTINDVNNENYDNGINKKLSYATVTKGENDVDKNSP-IEV 666
 QY 1153 LYONFNISDENGKWTISGLNPTFPADCCQLHEHTEGDKLVGNISMRIKTLDSGFTFR 1212
 Db 667 DVSNIN-----ELINFNANNRREVALLCNHRISIPQHOHTM 703
 QY 1213 SAVQKVFQDPNQGVLXNHD--NEXLHYODDWTYILSSKIEKRPEDYIFVYVYGRNDAMDGY 1271
 Db 704 SKIRKOIE---LYNEDIKRYKY-----LQHLKNSDKRKEIF----- 739
 QY 1272 GGAVVYTRSSVLPNSIIPLEKAKSIGRDFSTFIIDNTCGPPALVERLEKTEVEEGER 1331
 Db 739 -----SKVSTLDGTLRP-----NKKKEKMK 758
 QY 1332 IIVKEVEIEEVEKEVKEVGRTEMFLFOLAGFENELKODENFVRELSEKMEFLDE- 1391
 Db 759 -----EESECKRKL-----ITLIKVELLNNQMKVRDKNKTIALGTSKINYMPCR 802
 QY 1391 ---IKMPEASEVEKLFEGKALPIR 1409
 Db 803 ITYAFCKRFEIPIEKVFNRSLLK 826

RESULT 9
 670163
 hypothetical protein BB0512 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
 C:Accession: G70163
 R:Ritser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
 son, D.; Peterson, J.; Kewlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
 ; Bowman, C.; Gaitland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; M0ID:98065943
 A:Accession: G70163
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2166 <LFE>
 A:Cross-references: GB:AE001153; GB:AE000783; NID:g2688419; PID:g2688426; TIGR:BB0512
 A:Experimental source: Strain B31

Query Match 1.98; Score 144; DB 2; Length 2166;
 Best Local Similarity 16.88; Pred. No. 1.2;

Matches 263; Conservative 275; Mismatches 556; Indels 500; Gaps 71.

QY 44 RSNNGYFNFFLTSTKTSST--DSHCCKSOICSDTSFEELIQFDLRGM---TLI 98
 Db 482 RENNNSNLNDINKATYALTESLSSSSKFNOMESKYSFD---KLTAQMDFSLM 537
 QY 99 LKOKROPIQALIVCTVIVPRVAVDAKLTACCLKCRLELAKCIANSCAANVAC 158
 Db 538 YGKEFELISQ-----EATNNVQEPQDLNKLKNE----- 567
 QY 159 LQTCNNRPDETEQIKCGDLFENSVDQFNECAVSRKCCVPRKSDGEPVDRNAVQN 218
 Db 567 IESEFVNMFEKTQETLK-----VD-ENTSLIN-----IKDELIGK----- 599
 QY 219 FNMDFSGKWYITSGLNPFTDAFCQLEHFMENDKLVGNLWLRKTLIDGGEFTTSRQVOT 278
 Db 599 -NIYEFDRRY-----DEVNIFVTQLEESKIQYKMGOMDSNLKNT----- 641
 QY 279 FVODPDLGALYHNDNEFLH---YODWYIISQIENKPDYTFVYVYRGNDAMDGYG 334
 Db 641 -----SQINKNEFLSLIOQKDGIELSESVFNDLSDH-----QKALDMHGS 686
 QY 335 ---SVYTSRPL-----PESTIPNLQRAKSVGRFNNFTTDSGCPPLVERLE-- 385
 Db 687 WKHELIALKSLDILKVSSEDLSSATLKIESLEKDVN-----DMEYV 730
 QY 385 ---KTAEEGKLLIKKAVE-----IEEVEKEVEKVRDTEMTLFORLLEGF 427
 Db 731 LKRTGDI--ESLIEKYEKELKMSYQSDEALIGKEFINQETELIDKSVFMLEDLNKKF 789
 QY 428 KELQDEENFY-----RELKSEKEIINLOMEATE--VERLFGRALPIKILMA 475
 Db 790 ---DDKNMFYIKIEBCDYKLDKFIKIESEDIINFKSDLINEIES-----KLOIV 836
 QY 476 LAPH-----NFLANHETIKYKYKSLPGHKFSGWEDYEGSIYV--AK 518
 Db 837 SNKSDNOKOITDDELDRISKDILNRKDSINNEYDSKLS-----DMOSKINETIVK 889
 QY 519 ICSSRRIP-----RYFRKSPRIQCGILDRGQOLFESHGK----- 552
 Db 890 LSSGRVDLDLISEVTTKTELKFSIES-----LESYLEKIDEFNNQCAIISDLELQ 943
 QY 552 ---HNSPAHSINONVPRG-----NSGCKPRDYALVWVWKKWOPAK-TAIVAFITLAV 601
 Db 944 DINMHNKRETRLEENLSKRFPAAYLNNSBEYKVEVDSLDQKRDIASFOANDITLDSI 1003
 QY 602 ASKADAVDAKLTCTCLKCRLELAKCSN-PACAAVAC----- 641
 Db 1004 NYAFENDIN-----KEINCKINEVISNRYGSEINSSKLENEIHEIENLSRLLDR 1054
 QY 641 -----LQTCNNRPDETEQI-----KCGDLPE-----NSVYDFNCAVSR-K 677
 Db 1055 ISLSKGDENLQKLKESFVSKYQVEKFLKYDLDDEAKINLKVLEIDQYKSRLE 1114
 QY 678 KCVPRKSOVGDGPVDPSEVLQKDMKDFGSKWFTITGLNPTFDACDLHEFTHEENKL 737
 Db 1115 EADIRRTI-----DNDIM---QAKEREGE---ITNELKNNIESKSEPLNLYKERK 1162
 QY 738 V-GNLSMFIPTPDGFFTR--AVQKRVODPKYPGILYHNDNEVLLYQDDWYIISKRYEN 794
 Db 1163 IESNFEERYST---FLIESEGAISK-IRDELYKTLTSDEN-----LQIATISE 1206
 QY 795 SPEDYIIVYKGRN-----DAMDYGGSVLYTRSAVLPESITPELOTAQAK 841
 Db 1207 MDQNFELIORSKDIIEFEKLODKIKDQY-GFINSQGEIKAGVEENIKHNFQVCIKY 1265
 QY 842 GRDNTFIKIDNTCGPEPLVERLE--KVEEGERTIIKEV--EETEEVEKVRD---KE 894
 Db 1266 ---NLI--DDLIVKTEENIHKRISLSISSTFDSIEKLNNDVSCIRKIANDENLKY 1320
 QY 895 VTLFSLKEGFELODEENFLRELSEKMEVYLLGKMEALVEYKELFGRALPIRKILMAVA 954
 Db 1321 IELERECNQGOLNENKIKALIDNALISQYDGLKRYADMDYDESERL--NSYIATL 1378

QY 955 THCTSPCHDRIRFSSSDGIGRLGTRKING-----TFLKTLIPPIQ 998
 Db 1379 SEEFSSNKKMT--FELESOLKNLKNLESDDLNNVEADVIRLKEESHVNSHSLKEEDF 1436
 QY 999 SADLRTGGRSSRPLSAFSGSGCIDYVPLPSKNEKELAPLLKLVGLACAFILV 1058
 Db 1437 FKDLKIRGEELKYLENFASINDKIONLEYDISKN----- 1473
 QY 1059 PSADAVALKACALLGCGRIELAKCIANPACAAVACLOTGNNRPDETEQIKCGDLPE 1118
 Db 1473 -----LEKELLQSFRLIDIEQKMKDKENFLDFTKESSKKDMOSEIA--LME 1521
 QY 1119 NSV---VDENECAVSRKCCVPR-----KSDGEPAPDPVLYQVONISDPF---NG 1164
 Db 1522 TNIIGKVEFVDEVNKKQSIIDSWPLINKDVKDMQEKSYSTIKRINLAELGIKSPEND 1581
 QY 1165 KWYITSLNTPFAPDCQLEHFTTEGDN-----KLGNISWRKIKLIDSGFFT 1211
 Db 1582 IFNVKIGLESFQGFETKAEIIFSMQNEAKKIQSVHDFKNIGE-SLNLKVLDEKFEV 1640
 QY 1212 RSAVQKRVODPNO--PGVLYHNDNEVLYHOD-----WYLSKRIENKPE 1254
 Db 1641 DFLEKIDKVKRKTEDILIQAEVKFLQOKDLEDKLFEINOKLEHFTTSLSSNLNPKVR 1700
 QY 1255 DYIFVYVYRGNDAMDYGCAVYTRSSV-----LPNSITPELEKAASIGRD 1301
 Db 1701 EMVDYISSDK-----ESFQGLINKINSEFSKISLYRNINFTSIENYNSFSKISD 1756
 QY 1302 FSTFRTINTGCPPALVERIEKTEVEGEERIYKEVEELEEEVEKEVEKVGRTENTLFO- 1361
 Db 1757 LG-----LLEDEKLKSLKHS-----TSELETIKSGLQBDIK---FEV 1791
 QY 1361 RLAEGFNKLQDEENFRELKSKDEM-----EFLDEIR 1392
 Db 1792 EFKKHKKELLKEVDNNLLESKILNCQVQNFKFTSEIK 1830

RESULT 10
 B71603
 RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 13-Nov-1998
 C:Accession: B71603
 R:Gardner, M.J.; Tetelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2, sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600
 A:Accession: B71603
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1558 <GAR>
 A:Cross-references: GB:AE001424; GB:AE001362; NID:93845307; PID:93845309; TIGR:PFB091
 A:Experimental source: clone 3D7
 A:Genes: PFB0915w

Query Match 1.9%; Score 143.5; DB 2; Length 1558;
 Best Local Similarity 17.9%; Pred. No. 0.77;
 Matches 206; Conservative 153; Mismatches 377; Indels 413; Gaps 50;

QY 347 SIINPQK-AAKSVGDEN-----NFTTDSGCPPLVERLE-----RLKTAEBG 390
 Db 626 TLEENVEITTAESVTFNSNILEIOENTITNDTI---EKLELHEVNSAALENTQOSE 682
 QY 391 EKLKLEAVEIEEVEKEV-----EKVADTEMTLFORLKGFELOD---DENFVRL 441
 Db 683 EK---KEVIDVLEEVKEVATTLIFVQAEESASITTEIFENLEENAVESNENVAENL 739
 QY 442 SKREKEIINLOMEATEVEKELFGRALPIRKILMAVPHNSFLANHETIKYVYKSLPGHK 501

```

Db 740 EKLNELVNTVLKVEETVEISGESLENNEMOKAF-----FSEIDNNGKIOENLITG-- 793
QY 502 RFSWGWEDYFGSIYAKICSSRIKPRFKRSPICGGLDSRGLOJFSH-----KHNLSPA 557
Db 793 -----MERSITISIVQSE--KVDLNNVSSILDN--IENKKEILLKENISST 840
QY 558 HSNONVPGK-NSGCKFPKDVALVMWEMKQFAKTAIVAFILSVASKADAVDAKCTCTC 616
Db 841 EGOVEITVEHVOYVYDVDPAAK-----DOF-----LGLINEAGG----- 878
QY 617 LKKECRLKACISNPACANVACLOTCNNRPDETCQ---IKCGDLFNSVDEFNECA 673
Db 878 -LKEMFNLLEDVFKS-----ESDVIIVEEIKDEPVQKEVEKETVSIITEMENIVLDE-- 932
QY 674 VSRKRCVPRKSDVD--FPVPDPVLVQKEDPMKFGSKWFTIRGLNPTFDAPDCOLHEFH 731
Db 932 -----EKEDLDKMDVAEESIEISS-DKREFT-----Est 961
QY 733 TEENKLVNLSWRIRTPDGGFTTSRAVQFVQDPKRYGILYHNDNELLXQDDWYILSSK 791
Db 962 KKKKEDVSLYVEVQDND-----MDESVEKYLE-----LKNMEELMKDAVEINDITSK 1010
QY 792 -VENSPEDIYFYVYGRDANDMGVGSVLYTRSAVLPESI--IPELOTAOKVGRDENTF 848
Db 1011 LIEETO-----LNEVADLIKDMKLELEKA---LSEDSKEI 1046
QY 849 IKTDNCGPEPLVERLEKVEEGE--RTIIEVEEI-----EEVEKVRD-----K 893
Db 1047 IDAKD-----DLEKVEIEEHDITTLDEVELLKDVEEDKIEKSDLKLEDEDILK 1097
QY 894 EV---TLFSKLEPGKEIQDEENFLNLSKEEDVDLDGKMEATEVEKLEFGRAL----- 946
Db 1098 EYKEIKELSEILIEDKELTETILEKKEIKEDHEKKEPEAKEFKDLEADILKEVS 1157
QY 946 -----PIKRLMAVAHCTSPCHDRIRFSSDGIORLIGITRRKINGTFLKIL 994
Db 1158 SLVEEERKLEVEHLEKEVEHISGDAH-----IKG----- 1190
QY 995 PPIOSADLRTGGRRSRPLSAFSGSKGIFDIPLPSKNELKEITAPLLKLKVLCA 1094
Db 1190 --LEEDLEEVDDLKSIIDMLKGMELGMD--KESLEDTAKL----- 1231
QY 1055 FLIVPSADAVDAKTCACILKGCRIELAKCIANFACANVACLOTCN--NRPDTECOIK 1112
Db 1231 -----GERVESLKDVLSALGMEDEMKIRKKAQRKLEVLTK 1269
QY 1113 CGDLFNSVDEFNECAVSRKCVPRKSDLGEFPADPVSIVONENISDFNGKWTITSL 1172
Db 1270 -----EVEKEEPRKKITKK-----VRFDIKD-----K 1292
QY 1173 NPFTDAFCOLHEFHTEGDKLVNISMRIKIDSQFFTSAAVQKFOVQPNQGVLYNHD 1232
Db 1293 EPKKEIIVEEKKEDDED-----IEDEVED----- 1320
QY 1233 NEYLHQDDWYILSKIEKNPEYIFVYGRDANDMGVGAAYVTRSSVLPNSIILE 1292
Db 1320 -----IEEDKVEDDED-----E 1334
QY 1293 KAKSIGND-----FSTFRTDNTGCPBALVERIEKTEVEGERIYKEVE-----IE 1341
Db 1335 DIDEIDGEDKDEVIDLIYQKERRIEKYEKKKKLEKVEBVSGLKKNHVDVKKYQKID 1394
QY 1342 EEVEKEVEKV--GRTENTLFLORLAEGFNLKODE-----NEVRELKSEMEFLEDIKM 1393
Db 1395 KEVDEKSALESKNDVI-----NVLKONODFSKYNKVFYKKAFAAFISAVAA 1445
QY 1394 EASEVEKLE 1402
Db 1446 FASYVVGFF 1454

```

RESULT 11
S06006

```

myosin beta heavy chain, cardiac muscle - rat
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 13-Feb-1998
C:Accession: S06006; S67536; I67441
C:Krafft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
Nucleic Acids Res. 17, 7529-7530, 1989
A:Title: Complete nucleotide sequence of full length cDNA for rat beta cardiac myosin
A:Reference number: S06006; MUID:90016823
A:Accession: S06006
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-1935 <KRA>
A:Cross-references: EMBL:X15939; NID:956656; PID:956657
R:McNally, E.M.; Krafft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
J. Mol. Biol. 210, 665-671, 1989
A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Compari
A:Reference number: 597535; MUID:90133919
A:Accession: S07536
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-950, 7K 953-1935 <MCN>
R:Mandavil, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.
Eur. Heart J. 5, 181-191, 1984
A:Title: Cardiac myosin heavy chain isozymic transitions during development and under
A:Reference number: I53305; MUID:85179510
A:Accession: I67441
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1871-1935 <RES>
A:Cross-references: G: M32698; NID:9205598; PID:9205599
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methyl
F:88-766/Domain: myosin motor domain homology <MMD>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:548-585/Region: actin binding #status predicted
F:855-877/Region: actin binding #status predicted
F:839-1935/Domain: coiled coil #status predicted <COI>
F:839-1279/Region: S2
F:1280-1935/Region: light meromyosin
F:129/Modified site: H6/N6-tyrosine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:695/705/Active site: Cys #status predicted

```

Query Match 1.9%; Score 142.5; DB 1; Length 1935;
Best Local Similarity 16.7%; Pred. NO. 1.2; Indels 459; Gaps 56;
Matches 247; Conservative 218; Mismatches 558;

```

QY 166 PDETECOIKGDLFNSVVDQFNECAVSRKCVPRKSDVGEFPVDRN-----AVV 216
Db 375 PDGTE-----LADKSAIYLMGLNSADLKGCHPRKYNKNEVTFGVAQVQVAAVAGALA 428
QY 217 QNFNMKDSGKVYITSGLNP-----FAAFDQLEH-----FHEENKL 255
Db 429 KSVYERKEN--MMVTR--INATLETROPYFLGVLDIAGLELFDNSFQCLINTNEKL 485
QY 256 VGNLTWIKITLDGFTFRSAVO--TFVO-----D 282
Db 486 QQFNHHRFVLEDEYKKEGIEWTTFIDFGMDLOACDILEKMGINSILEECMPKPAD 545
QY 283 PDLGALY--NHNDNELLHYDDWYILSSQIEKNRPDIYFYVYGRND-----ANDGIGGSYI 337
Db 546 MTFKAKLYDNHLGKSNFQKPRNKGQEAH-----FSLIHAGTYDYNLIGW-----L 594
QY 338 YTRSEPTPESTLIPMLQAA--KSGRDNNFTITDNSGCEPPLVERLEKTAEGEKLKIK 396
Db 595 QKNKDPLNETVVGLYOKSSKLKLSNLFANT-----AGADAP--VDGKGRKAKKSSGFTV 647
QY 397 FAVEIEEVEVERVEKVRDTEMTLEFOLLLEGFKLODEENFVLESEKE----- 448
Db 648 SALH--RENLMKMLNLRST-----HPHFVRCIIPNETKSPGVMDNPL 688

```

QY 448 LNELMEAT-EVERLEFGALP-----IRKLMALAPSNFLANHEITIKYVGS 495
 DB 689 VMOHJRCNGVLEGICRCKGFPNRLTYCDFORRRIINPAIPFGGQIDRKAEMKLE 748
 QY 496 KLPGRKRSWGWEDYFGSIVVAKICSSRPIRYFRKSPRJCGLDLSGLOLFSHGKHL 555
 DB 749 LIDIDNOYKFGHTKVFYFAGLIGLEEMRDERLSRITRI--QASGVSJRNHEFKLE 806
 QY 556 PAHS---INONVPKNSGCKFPKOVAILMVEKMGCFKTAIVAFILISVASKADAV---- 609
 DB 807 RRDLSIIOWNIR-----AFMGVKNFMKLYFKIKPILKSAETEKANMKE 854
 QY 609 -----DAKTCTULKEBRLIETAKCISNPACANVACLOTCNNRPDEECOKGDLFE 662
 DB 855 EFGVKALKESEARKELEKMWLSLQF-----KNDLQLOVQEDONLADAEKRCOLIK 910
 QY 663 NSV-----VDEFNE-----CAVSRK---KCPKRSVSGFPVPDPSSVLYOKED 702
 DB 911 NKILEAKVEMTERLEDEEMNABELTAKRKLEDECSLEIKRIDIDLETLAKVEREK-- 969
 QY 703 MKDPSGWETTRGLNPTFADQCLHEFTEENKLVGNSWRIRTPDGGFTFRSAVQFV 762
 DB 969 -----HATENK-VKNLEEMAGLD-----ETIVKLT 993
 QY 763 QDPKPGIYNNHNDXYLLYODWYILSKVENSPEYIFVYKGRNDWDGYSGLYTR 822
 DB 994 KEKKALQEAHQALDLOAEEDKVNLTTRAKAKYLEQOV-----DDLEGSILDQK 1042
 QY 823 SA-----VLPESIIPELOTAQKVG-----RDF-----NTFKIDNTGCP 857
 DB 1043 KVMDELERAKKLEGLKLTQESIMDLENDKQDLERLKKRFEINAINARIEDQALGS 1102
 QY 858 E-----PVEYERLEKVEGERT-----IKVEEIEEVEK----- 890
 DB 1103 QLOKRIKELOARIELEEBL-EAERTARAKVEKLSRDSLELEETSELEAGAGATSQOI 1161
 QY 890 -----VDRKFEVTLFSKLEFGFKLODEENFIREL 919
 DB 1162 ENMKRREAFQKMRDLLEATLOHEATAALRKHADSVALEGEOIDNLOKVKOLEK 1221
 QY 920 SKEEMOVULGLMEATVEYKLEFGRALPIRKLMANVATGFTSPCHDRIRFSSDDIGRLG 979
 DB 1222 SEFKLELD-----VTSNMOIITKANANLEKM-----CRT--LEDOKNHRS-----KAE 1264
 QY 980 ITRKIRNGTFLKLIPQASDLRTTGGSSRP-----SAFRSGSKIPDIPIPSKNEL 1036
 DB 1265 EFORSVNDL-----TQRAKLOTENGELSROUDEKALISQLTRG-----RLTYTOOL 1312
 QY 1037 KELTAPLLKLTVGLVACFLVPSADAVDAIKTCACLLKGRIBLAKCIANPACANVAG 1096
 DB 1313 EDUKROLDEEVAKANALALQASARHODDLRBOYEETFAKALQRYLSK--ANSEVAG 1370
 QY 1097 LOTCNNRPDETECOIKCGDLFENSVDVEFNCAVSRKKCYPRKSDGGEFPADPSVLYON 1156
 DB 1371 WRT-----KETDAIQOTLELEPAKKALQRLQDAE----- 1403
 QY 1157 FNISDFNGKWTYTGCLNPTDADFQCLHEFTEENKLVGNSWRIRKTLDSGFTFRSAVQ 1216
 DB 1403 -----AVEAVNAKCSLEKTRHLQNELELDMD-----VERSNNA 1438
 QY 1217 KFYODPNQGVLYNHNDNEYHYODDWYILSKLEKPEYIFVYGRNDWDGYSGLVAV 1276
 DB 1439 AALADKRO-----RNFEDKILYEMKOKIYESOSELESOSKE-----ARSL 1477
 QY 1277 YTRSSVLPNSIIPLEF-----KAASIGHDFSTFRTONTGCPPALVERIEKTEVEGE 1330
 DB 1478 STELFKLKNAVEESLEHLETFKREKNLQDEISDLTEQJSGTSGKSHLEKIRKOLEK 1537
 QY 1331 RIYKVEEIEEVEKEVEKVGRTMTLQ-----RIABGFNELKODEENFVRELS-- 1382
 DB 1538 LELQSALEFEASLEHEGKILRAQLEFNQIAELIERKLAEDKDEMEQAKRNHLKVVDLS 1597
 QY 1382 -----KEEME-FLDEIKMEASEVEKLEFGKA 1405

Db 1598 QTSIDATFRSRNALKYKKMBGDLNEMETQLSHANMAAFA 1639
 RESULT 12
 A:71928
 C:ag island protein - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 C:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence revision 12-Feb-1999 #text change 12-Feb-1999
 C:Accession: A71928
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
 ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
 A:Reference number: A71800; M01D:99120557
 A:Accession: A71928
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1819 <ARN>
 A:Cross-references: GB:AE001481; GB:AE001439; NID:94155005; PID:94155007
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: orf13/14
 Query Match 1.9%; Score 142.5; DB 2; Length 1819;
 Best Local Similarity 19.0%; Pred. No. 1.1;
 Matches 267; Conservative 199; Mismatches 524; Indels 417; Gaps 68;
 QY 178 LFENSVDQFNECAVSRKKCYPRKSDVGEPPVDPNNAVONFMKDPFSGKWTSGINPT 237
 DB 244 LFSRSIFHY-----VPLEDKSSRFS-KDRNLYVND----- 274
 QY 238 FDAFDQCLHEFHEMENDKLVGNTWIRIKTLGDFGFTFRSAVQTFYQDDDLGALYNHNEFL 297
 DB 274 -----EIQIRQ-----EYNRLKERNKGMIDKNF-----INDDEN--RTLYN----- 312
 QY 298 HYODDWYILSQIENKDDYIFVYGRNDWDGYSVYTRSPILPESILPNLOKAAK 357
 DB 312 -----YLNIALEEDK--NPLRAFYECISN-----GNN--YEBCIKLIND--KRIQDDQK 354
 QY 358 SVGRDNFNFTTNSGCPPEPLVERLEKTAEEGK--LLIKRAVEL-----EEBV 405
 DB 355 KTLBAYNDCL--KNKTEBERIKCLDLKDEKLSKLNDQKVOVALDCLKNATDER 411
 QY 406 EKEVEKRDTEM--TLQRLLEGFKLQODEENFVRELSKEEK--EILNELMEATEVEKLE 463
 DB 412 KECLKLINDEIREEKREKELELOKELOEYKDCIKNAKTEAEKNECKLSKEA--IEHLK 469
 QY 464 GRALPIKRLMALAPSNFLANHEITIKYVGSKLPGHKRFSWGWEDYFGSIVVAKICSSR 523
 DB 470 QQALDDCKKNAKTDEBKKECLKN--IPDLOKELLADM-----SVKAYKDCVSR 515
 QY 524 RIRPIYFRK-----SPRICGDSRGLQLFSGKHLNLPASHINONVPR-----GNSG 570
 DB 516 ARNEKEKOECEKLLTPPAKKKLENOALDOIKNAKTDEBKRECL--KNLPKDLQSDILAKES 574
 QY 571 CRFPKDVN-----LWMEKMGOFKATAIYAILISVASKADAVDAIKTCTC 616
 DB 575 LKYYKCAQAKTEAEKKECEKLLTP-----AKKLEEAKEKSVKAYLLDCVSAKT--EA 628
 QY 617 LKKEC-----RLBLAKCISNPACANVACLOTCNNRPDTEGO-----IKGGLFE 662
 DB 629 EKKECEKLLTPPAKKKLEBEAK-----KSVRAVLDCVSAKNAEKECEKLLTPPAKKLE 684
 QY 663 NSVVDVEFNCAVSRKKCYPRKSDVGEPPVDPSSVLYVOKFMDKDFSGKWTTRGINPTFDA 722
 DB 685 NQALD-----CLKNATDERRECKLDP-----KDLQKVLAKESVRYLDC 727
 QY 723 FDCQCLHEFTEENKLVGNSWRIRTPDGGFTFRSAVQKFPYODKRYPIILY----- 773
 DB 728 VSKARNEAEKRECE-----KLLTPE-----AKKLEEAKKSVKAYKDCVLRARNE 772

QY 773 --NHDNEVLLYQDDWYIISKVENSPEDYIFVYKGRNDAMDYGSGVITYRSVAVLPESI 830
 DB 773 KEKOCCEKLLTPPEARKLLEES--KRSVAYALDCVSKAKNEERKECEKLLTPPEARKLLEEA 831
 QY 831 IPELTAQOKRGDENTFIKTNDTCGPEPLVERLEKKEVEGERT-----IIEVEEL 883
 DB 832 KRSVAYADCSVRANE--KEKOC--EKLLTPPEARKLLEESKRSVAYALDCVSKAKNEA 887
 QY 884 E-EVEEYKRDENVTLFSLFEGFKEL-----ORDEENFLRELSEK-----EM 924
 DB 888 EKECEKLLTPPEA--RKLLLEAKESVAYKDCVSRANEKECEKLLTPPEARKLLEEN 944
 QY 925 DVLDDGKMEATEVE--KLGRALP--IRKLMAVATHCFTSPCHDRIRFSSDGIIGRLI 980
 DB 945 QALDDCKNAKTAERKRCVKDLPRLOKKVLA-----KESVAYALDCVSKAKNEA 994
 QY 981 TTRRINGTFLKILPPIOSADIRTTGSSRPLSAFRSGSKGIFDIYPLPSKVELKELT 1040
 DB 995 ERECE--KLTP--EARKLLEAKESVAYKDCVSRANE--KEKOCCEKLL 1041
 QY 1041 APLLLKLV-----GYLA-----CAFLVPSA-----DAVALKTC-- 1071
 DB 1042 TPEARLLEOEYKSVAYALDCVSRANEKECEKLLTPPEARKLLENOALDCKNAKT 1101
 QY 1071 ACLKGCRIEL-----AKCIANPACANVACLOTGNRPDETECO-----IKCGDLFENS- 1121
 DB 1102 EAERKCVMDLPKDLQKVLAKESVAYALDCVSRANEKECEKLLTPPEARKLLEESK 1161
 QY 1121 -VDENECAVSKRCVPRKSDGFEPPADPSVLYONENISDFNGKWYITSGLNPTEFAF 1179
 DB 1162 KSVAYALDC--VSKAKNEAEKCEKLLTPPEARKLLE-----AKESVAY 1205
 QY 1180 -DCQHEFHTEGDNKLVGNISMRITKLDGCFTRSAVOKFVODPNQGVLYNHDEYLYH 1238
 DB 1206 KDC-----VSARNEKEKQ-----ECEKLLT 1226
 QY 1239 QDDWYIISKVENSPEDYIFVYKGRNDAMDYGAVYTRSSVLPNSIPE-----LEKA 1294
 DB 1227 PEARLLEOEYKSVAYALDCVSRANE-----KEKOCCEKLLTPPEARKLLENO 1275
 QY 1295 AKSIGRPFSTFIRNDTCGPEPALVERIEKTEVEGERIIVKEVEEIEEVEKEVEKVT 1354
 DB 1276 RQOKDKAIKDCIK--NADPNDRALIMCGLDGLSDEKLT--KYLQEAERKAVLDCLKTART 1331
 QY 1355 E-----MTLFORLAEGF--NELKODEENFVRELSEK-----MEFLDE----- 1391
 DB 1332 DEERKCCONTLYSDLIQETIONKKAONKON--QLSKTERLHQASECDLNDLDDPTDQAIHQ 1388
 QY 1391 -----IKMEASEVEKLF 1402
 DB 1389 CLEGLDSEERALLIGIKRAODEVDNITY 1415
 RESULT 13
 C70319
 nltirite reductase (NAD(P)H) large subunit - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
 C:Accession: C70319
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; OY
 V
 Nature 397, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666
 A:Accession: C70319
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1002 <ACF>
 A:Cross-references: GB:AE000678; NID:g2983921; PTD:g2982926; GB:AE000657
 A:Experimental source: strain VF5
 C:Genetics:
 A:Gene: nltB

Query Match 1.9%; Score 141; DB 2; Length 1002;
 Best Local Similarity 20.1%; Pred. No. 0.59;
 Matches 192; Conservative 148; Mismatches 314; Indels 300; Gaps 54;
 QY 289 LYNDNEFLHYQDDWYIISL-----QIENKPPDDYIFVYKGRNDAMDYGSGVITYR 340
 DB 90 LYTDDGEFYR--DKALITAGSKPFIPIPSIKGVNKGVEFT--YRTADVPEILLDYARVSKR 146
 QY 341 SPLPESIIIP--NIQKAAKSVGRDNNFTITDNCGPEPLVERLEKTAEE-----G 390
 DB 147 AVVIGGGLIGLEVTKALRDIGLEFVLHIDLTL-----MEQDITASELLRKLDLMDG 200
 QY 391 EKLITKEVEIEEVEV--EKEVEKVR-----DTEMIFQRLLEGKELQODENEFRELSE 444
 DB 201 IKVLK--KYTEELGKRAEGVRFSDGELLADFIYATG--IRPNVEGVNSGLKV 254
 QY 445 EKEIL--NELOMEATEV-----EKLGRALPFRKLALAPHSNLANHETIKY 492
 DB 255 NKGIYVNDYLETASDIYAVECEIEHRGKTYGLVAPIME--QVYCAHN--VHGNEKKY 310
 QY 493 VSS-----KLGHKRFSGW--WEDYFGSIYVAKICSSRRIPRYFRKS-----PRIC 537
 DB 311 TGLTYAMLKAVGVNLFESAGEINEKDGDEVYAFIDNGRSL--YRKAVINNNKIVGTHLY 367
 QY 538 GLDSRG-----IOLFSHGKHNLS--PAHSINQVNPKNKSGCKFKDYALAMWEKMGQFAK 590
 DB 368 G-DVRGNNNYLLDLKSGK--DISERPYFLIKHILPKGTGVEELKD----- 412
 QY 591 TAIVAFITLSVASKADAVALKTCITCLKEORLEAKCISMPACANVACLOTGNRPDE 650
 DB 412 -----NDIVCNNAVTKG--ELVKQIRE-----GC--KILEEIQER 443
 QY 651 TECQIKCGDLFENSVDDEFNECAVSKRCVPRKSDVGDGPVDPDSVLYOKDKAKDFSGKW 710
 DB 444 TKASTGCGSCITE--LVEELLKHVYKEK--PKR-----VNKIEV----- 478
 QY 711 FITRGINPTFADFQOLHEFHT-----EENKLVGNLSW-----RIRT 747
 DB 478 -IKKEIHPFLFELFKRLEKRYFSGELENIPEEDRDV--RLKMGITRYKATGYEVNRRIV 535
 QY 748 PDGCF--FTRSAVOKFVODPRYPGILYNHDEYLYQDDWYIISKVENSPEDYIFVYK 806
 DB 536 PNGRLSTEOKAKVYSHISEKRCRGEVETLSROOL--QIRMI-----KLKDLPEIL----- 583
 QY 807 RNDAMDYGSGVITY-----RSNAVLPESIIPELQIAAO-----KVG 842
 DB 583 --EALNRYGLSTLOTGMDNRYNTGDLPLGLADESLIDLRLRSQETITNIFLCKKKYADIP 640
 QY 843 RDNFTFIKTDMCGPEPLVERL-----EKKVEEGERTII-----KEVEELE 884
 DB 641 RKLNAV--LGSQDQCNALFNDVCFYLSKDGKLGNNLYLGKISSGGPKRAIDMD 695
 QY 885 EEVERKRDKEV--TLFSKLEFGKELQDEENFLRELSEKENDVLDGLKMEATEVEKLF 941
 DB 696 MEVEPEYEVYVATL--DIYSTFGNRENSKRNLYFLQEW-----GVERRELEKRL 748
 QY 942 GAALPIRLKLVAVATHCFTSPCHDRIRFSSDGIIGRLITRRKIRNGTFLKILPPI----- 998
 DB 749 YKAIP-----SKGKDLVNTGTEGREGIINR-----NGYACVAVPAGKIK 789
 QY 998 -----QSADL--RTTGRSSRPLSAFRSGSKGIFDIYPLPSKN-----ELKELT 1040
 DB 790 AADFQALFELAKRYGSRRLR--LSVYQNY-----INPIEENINPLUGEIEFEKSTVS 842
 QY 1041 APLLLKLV-----GYLACAFILVPSADAVALKTCACLLKGRIRIELAKCIANPACA 1091
 DB 843 SPTMTMLIACAGSDTSCFGVIPKNS--DAVRVAKYLSERKLDLIPVMMHSACA 894
 RESULT 14
 JCG009

surface-located membrane protein lmp3 precursor - Mycoplasma hominis (SGC3)
 C:Species: Mycoplasma hominis
 C>Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 05-Dec-1998
 C:Accession: J05009
 R:Labelled, S.A.; Jensen, L.T.; Brock, B.; Birkeland, S.; Christiansen, G.
 J. Bacteriol. 178, 2775-2784, 1996
 A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp gene system
 A:Reference number: J05009; MUID:96213016
 A:Accession: J05009
 A:Molecule type: DNA
 A:Residues: 1-1302 <LAD>
 A:Cross-references: EMBL:X95601; NID:91197335; PID:e222344; PID:g1197336
 C:Genetics:
 A:Gene: lmp3
 A:Genetic code: SGC3
 C:Superfamily: surface-located membrane protein lmp3; tetratricopeptide repeat homology
 C:Keywords: duplication; membrane protein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1302/Product: surface-located membrane protein lmp3 #status predicted <MAT>
 F:957-992/Domain: tetratricopeptide repeat homology <TT1>
 F:993-1026/Domain: tetratricopeptide repeat homology <TT2>
 F:1029-1120/Domain: tetratricopeptide repeat homology <TT3>
 F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match Best Local Similarity 1.9%; Score 140.5; DB 2; Length 1302;
 Matches 173; Conservative 123; Mismatches 312; Indels 315; Gaps 40;

```

QY 145 KCIANPS-----CAVNACTLOT-----CNNRDETECOIK--CGLEF--NSVVDQENE 189
Db 445 KSVTNSSNKSIIANALLQALADAKADQVDEANKSTKEQLNALIDKATLTLLPOLIND 504
QY 190 CAVSRKCKVPRKSVGFPVDRNAVQENFMKDFSGKWTISGLNPTPDAPCCOLHEFH 249
Db 505 ---NDSEIVAKESLNAETNANKAVONDNMSQSAK---SSLDKATKIQNLTFEN 557
QY 250 MENDKLGNLTWRIKTLIDGGEFTRSVQTFVQDPDLGALVHNDNFLLHYQDDWYILSSQ 309
Db 558 KDKDA-----KFKLEQ---TRKIDNLFLLD-----VKNNPN-----VATLVKD 594
QY 310 IENKPPDYIVYIRGRNDAMDVGSGSVYTRSPILPESIIIPN-----LQ 353
Db 595 LTNAKDKRSV-----TKSSNKSSEIIAANDDELKQALDAAKAKQDID 636
QY 354 KAAKSVGRDFENFTDNCSCGPEPLVERLEKTAEGEKLIKAVETEEVEKEVKVR 413
Db 637 EANKSIKEDLSBITNANO-----LNLKL---VDSKDIDQAKATELSQFISASQELN 686
QY 414 -----DTEMTLFGRLLEGEKELQODEENFVRELSEKEKEIINELQMENTEV 459
Db 687 LNNPTSMQSAKESLDAKVEITRKLETF---NKDKVAFKLEKTRDI----- 733
QY 460 EXLFGRALPIRKLRMALPHSNELANHETIKYVGSKLPGHKRPSWGMEDYFGSIYVAKI 519
Db 733 -----DEFINTNKNPNY-----STLISEL 752
QY 520 CSSRPIPRFPRKSPRICGLDSRGQLFSGHKNLSPAHN--INONVKKNSGCKPEKQVA 578
Db 753 TSKR-----DSKNSVITNSNKSXSDIETANTTELKQALAKANTDRAQADNLA 796
QY 579 LMVMEKWCQFAKIALVAFILSVASKADAVDALACTQCLKEGRLELAKISNPAA--N 637
Db 797 KSTKEQLNNSISSA--NTLAKLTDKDNTLQOAKT-----ELEKEVQK--ANQAVASNN 846
QY 638 VACLQTCNNRDETECOI--KCGDLFENSVDNECAVSRKCKVPRKSVGDP-----PV 691
Db 847 TASMQSAKSLADAVTELTKLETFNKDKVKELEQTRK-----DIDEFINTNKN 899
QY 692 PDPVAVLQKPEMKDFSGKWEITRGLNPT-----PFAFDC----- 726
Db 900 PDYSTLISELTSKRD--KNSITNSNKSXSDIETANTTELKQALAKANTDRAQADNLA 958

```

```

QY 726 OLHEPHEENKLYGNLSKRITPDGGEFTTSVQKEVQ-----D 764
Db 950 QUNKISSANTLLAKLTDKDNTLQO---AKTELEKEVKANQAVASNNNTASMQSAKSSLD 1015
QY 765 PRYPGI-----LYNHDN-----EYLLYQDDWYILSSKVE-----NSPEYIIV 802
Db 1016 AKVEITETKLETFNKDKVAFRELEQTRKIDIDEFINNTNKNPNYSTLISELTSKRD--- 1072
QY 803 YKGRNDAMQVGSVLYTSAVLPESII-----DELQTAQKVGDFENFTIKTDNTCG 856
Db 1072 ---SKNSTTSSNKSXSDIETANTTELKQALAKANTDRAQADNLA 1127
QY 857 PEPLVERLEKKEVEGERITIKVEEIEEVEKVR-----DKEVTLF 898
Db 1127 -----LAKLTDKDNTLQOAKTELEKEVKQANQAVASNNNTASMQSAKSSLDTRKVEI 1178
QY 899 SKLPEGRKEIQRODEENFLRELK 921
Db 1179 TKLETF---NKDKRAKFMELKK 1198

```

RESULT 15

F64508

hypothetical protein MJ1672 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997

C:Accession: F64508

R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Weinstock, R.G.; Merrick, J.M.; Glodex,

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999

A:Accession: F64508

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-800 <BDU>

A:Cross-references: GB:U67607; GB:L77117; NID:g1592244; PID:g1500575; TIGR:MJ1672

A:Map position: REV156207-1653805

Query Match Best Local Similarity 1.9%; Score 140; DB 2; Length 800;
 Matches 180; Conservative 96; Mismatches 228; Indels 394; Gaps 45;

```

QY 650 ETEQIKCGDLFENSVDNECAVSRKCKVPRKSVGDFPVDPDSV--LVQKF--DKMD 705
Db 119 ETVC---IGELTENLYKNGFK---YSLKPLVNSDALFTDKPRPNENYKELFSKFEDEIKD 172
QY 706 FSG-----KFL-----TRGINTPFDADF-----COLHEFH 731
Db 173 FKGVSEELYOLMQKTYWCIPSVIMKKRAGSLKGLPVSLEFDSKTKTCAIACCLQXMY 232
QY 732 TEENKLVGNLSWRIRPDGGEFTRSVQKFPVODPKPGILVHNDY-----LLY----- 782
Db 233 VKENK-----KKNKYAKEYIDDKLEKLEFNNDGMWKEIFSLIHGDL 275
QY 782 ---QDDWYILSSKVENSPEDYIEVYKGRNDAMD-----GYGGS 817
Db 276 GIDQVFETITF-----YATKSLKGRSFYLDPLETFEYAKYICKELMLPTINILFYGG 328
QY 818 VLYTRSAVLPESTIPELQTAQKVGDFNTF-----IKDNTCGPEPLVERLEKKEVEGE 873
Db 329 HFYILSKYVDENLNLKEINEV--LFNMFRTKIYITIAEYGVTPNDFKKSSEK--ESKE 385
QY 874 RT--IIKEVEELIEEVEKVRDEVLFLSKLFEQ--FKELQRODEEN----- 915
Db 386 KTWGFAKKWKEVS---EKVEKKLRREFYKLEGLFEYPRNRSNKKVCYICRNEFDKNGKY 442
QY 915 FLRE--LSKEE-----MDVLDGKMEAEVEKELGRALPIRKLMNAVATHCPTSP 961

```

```

Db 443 AIRENESKSERIODYCAFEVALFDILKNFOMEXT---IKFNKAYPI----- 486
QY 962 CHDRIRFFSSDDIGRLGTRKRINGTFLKILPIQSDLRRTGGRSSRPLSAFRSGFS 1021
Db 486 -----IHLTKNDN-----LSLOREFS 503
QY 1022 KGIFFDIPLPSKNELKELAPLLKLVGLACAFILVPSADAVDALKTCCCLKGRLEL 1081
Db 504 -----FLTV---KAIEKLESKFVLSLSD----- 523
QY 1082 AKCIANPACANAVACLOTNNRPDETECOIKGSDLEFENSVDENFNCAYSRKKCVPRKSD 1141
Db 523 -----ENYFLKEYKLPDSDGELIIPYKIM 546
QY 1142 LGEFPAPDPSVLVON---FNISDFNG---KWTITSG-----LNPTFADFCDQLHEFHTE 1189
Db 547 ALAFP-----IENETEKRIIDFDGLAEKAFERTIGTRKIGIKMDVD---NLGEIFTT 596
QY 1190 GDNKLVGN---ISWRIKTLDSG---FETRSAYQKFOVDPNPGVLYNHNDNEYLHYODDWY 1243
Db 597 G---LGNDAITIS-RMSTLSSMLTFFTYI-----PHLIKNEFE----- 633
QY 1244 ILSKTIENKPEDYIFVYIRGRND---AMDYIGGAVVYTRSSVLPNSIIPLEKAKSI 1298
Db 633 -VNGK-KYFKKNDIYLVYAGGDDTLIVGAWDA-----VWELAKRI 670
QY 1299 GRDFSTFIRDTNC-GPEPAL-----VERIEKTVEGEIRIIVKEVEITE 1341
Db 671 RGDFFKFFV---CYNPYITLSAGIVFVNPKEFFKAVNNAEPELENGKNYIIEDEDETE 725
QY 1342 EEEVEKEVEKVGRTM-----TLFORLAGFNFELKODENFYRELKREME 1386
Db 726 KKYDKNALIVFNCPPMMWDEVEYNEVCWTKRKYLEGINKEMVELESIVKKFNEDDLE 783

```

Search completed: October 14, 1999, 03:19:10
 Job time: 2196 sec


```

Db 641 QULERANFEQESTLSSENNDLRTKLKLEESNKLKROEDVDSLEKNIOQLK----- 696
Qy 496 KLBGHRKFSWEDYGSIVAKICSSRRIRPRFRKSPRICCGDSHGLD---SHGK 552
Db 696 -----ED-----LRKSEELRFSK-----LEAKNLEVIDNKGRK 726
Qy 553 NLSPA-----HSI-----NONVPKNSGCKFPKDVALWMEKMGQAKTAIVAIFLISA 602
Db 727 ETLERQNDLHSSLSDAKNTNAIISSELTKSSEDV-----KRLTAVELLTD 774
Qy 603 SKA-----DAVDALKTCTCLKECRLELAKCISNPACANVACLOTCNNRPDETCQK 656
Db 775 SKAMQSTSLVNSIQSISNLYHELRODHVN-----MOSQNTLTLESSEK 821
Qy 657 --CGDLFEN--SVYDEF-----NECAVSRKCVPRK-----SDVGFVPVPP 694
Db 822 TDCELTQOQNTLLDNQKLMKHVNOESVSEKLENGSLSDLNKSSLVVAISDND 881
Qy 695 SVLVQKFDK-KDESGKWFITRGILNPTDAFDQCLHEFTE-----NKLVGNLSWRIR 746
Db 882 QILTQELSKNSYDLEOESQOLNSGLKSLAEKQILHTENEHLIRDLKLGK 938
Qy 747 TPDSGFFRSAVQFVODPKYPIGLYHNDHYLLYODDWIILSSKYVNSPEDEYFYVYK 806
Db 938 -----IESKSSDL---GKLTARQEE---ISNKEENMSQSAITSYKS 976
Qy 807 RNDAMDYGGSVLYTRSAVLPESILPELOTAQKVRDENTFKTDNCGPPLVERLE 866
Db 977 KLDE-----TLRSKSK-LADIEHLKMKSEVEVERNALIASN-----ERIM 1017
Qy 867 KAYE-EGERTIIKEVEIEEVEKVRNDEVTLSKL-----FEGR-----KELORD 911
Db 1018 DDLKNNSE-----NIASLOTEIEKKRAENDLOSKLSVSEYENMLLISSQTNKSLD 1071
Qy 912 EENFRLSKEMDVIDGLKMEATEVEKL---FGR-----ALPIKMAVATHC 957
Db 1072 KTNOLKTIENKOVKLDKQORVLELELTISKYKLGEMNAIKDELLARKK----- 1125
Qy 958 FTSPCHDRIRFESSD-----DGIGRL-----GIFRKRNGTFLIKLPIOS--AD 1001
Db 1125 -SKOHDLKANFYDDLKESDLEQJLTKNKNELVLSLEBSNNNALVEESDLANRLSD 1183
Qy 1002 LRTTGGSSPPLSAFSGSGSKGIFDIPLPSKNEKELTAPLLKVLGACAFIYPSA 1061
Db 1184 MKRSLSDSDNVISVIRSD-----LVRVN 1206
Qy 1062 DAVDALKTACALCKGRIELAKCIANPACANVACLOTCNNRPDETCQKGLFENSV 1121
Db 1207 DELDITLK-----KDKDSLSTQYSEVCOEDRDLDELKGC----- 1241
Qy 1122 VDEFNCAVS-RKCVPRKSDGEPFAPDPVLYON--FNISDFNGKMYIT--SGLNPTFD 1177
Db 1241 EESFNKAYAVSLRELCTKSLIDV-----PVSELTDNPFNMGNSLSRLVLSLENYLD 1295
Qy 1178 APDCQULHEHTGDKNLVGNISWRIKTIDSGFTFTRSAVQKFDVDPNPGVLYNHDNEYLH 1237
Db 1296 AFN-QVNEKMELDN-----RLTTDAEF-----TKYVADL-----EKLQH 1330
Qy 1238 YODDWIILSSKIENKPEDEYFYVYGRNDAMDGVGAVYTRSSVLSNLSIPELEKAKS 1297
Db 1331 EHDWMI-----ORG-----DLEKALKD 1348
Qy 1298 IGRDFTPLTDNCGPPLVERLEKTEVEGERIYVEVEIEEVE----- 1346
Db 1349 SEKNF-----LRKAEMTENIH-SLEGGKETKKEIHELSSRLIEDNOLATNKIKNO 1398
Qy 1346 -----KEV-----EKVGRTEMTLFORLASEFNELKODEENFVAEISKEDEMEFL--DEIKMEA 1395
Db 1399 IDHLNOEIRLEADVLKRESLIISLEESLSNOROKESSLID--AKNLEHMLDTSKXNS 1456
Qy 1396 SEVEKL 1401
Db 1457 SLMEKI 1462

```

```

RESULT
2
RBP1_PLA VB STANDARD: PRT: 2869 AA.
ID RBP1_PLA VB
AC 000798;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE RETICULOCTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1.
OS PLASMODIUM VIVAX (STRAIN BELEM).
OC EUKARYOTA, ALVEOLATA, APICOMPLEXA, HAEMOSPORIDIA, PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92315338.
RA GALINSKI M.R., MEDINA C.C., INGRAYALLO P., BARNWELL J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL CELL 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M88097; GI60626;
DR HSSP: P36956; IAM9.
KW SIGNAL. RECEPTOR. SIGNAL: TRANSMEMBRANE.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 2869 RETICULOCTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SO SEQUENCE 2869 AA; 330213 MW; 2BD43E2C CRC32;

Query Match 2.1k; Score 155.5; DB 1; Length 2869;
Best Local Similarity 17.4k; Pred. No. 0.59;
Matches 283; Conservative 242; Mismatches 573; Indels 529; Gaps 76;

Qy 57 TSYKTSFSDSSHCKDRSQICSDTSFEET-----QRFDLKGMWLLLEKOWRQIOLA 110
Db 1150 TSNKNEILKSVKEVEDKLVYEQNEDYKKVKNPENKQLEALIGSSKIKE----- 1201
Qy 111 IYVOTFYFVAVVAVATLCTACALKECRLELAKCIANPSCANVACLOTCNNRPDETE 170
Db 1201 -----VINKHVSQMQLSTANTLK-----SNAKGKENGHDEJELN---KIT 1239
Qy 171 QOIKGDLFE--NSVDOFNCAVSRKCKVPRKSDGEPFVDPDRN--VYONFNM-KDF 224
Db 1240 GQMR--DIYEKTKIAELKEGYVNEKLDANEKANVY--PEPERNIGHVLERTIVEXDK 1296
Qy 225 SGKMYITSGNLPTFADFQCLHEF--HMENDKLVGMWIRIKTIDSGFTFTRSAVQTFOD 282
Db 1297 AGK--VVEENMSLKTREKLIQETSDSONELVYTSITHLDMAG----- 1341
Qy 283 PDLGALXNNDNEFLHQDWIILSSQIENKPPDYFYVYGRN--DAMDVGGSVLYTRS 341
Db 1341 -----YED-----VIRKNEDSIOLEKAKSLFTLD-----EM 1368
Qy 342 PTLPESTIIPMLQKAAK---SVGRDFNN-----FTTIDNS----- 373
Db 1369 KRLVQGVNMNLSAIOGNAGISKELNELKGVIELLISTVYSILEYVKKNSSESVRSOL 1428

```

```

QY      1268 -----WNYGGAVVTYTRSSV--LPN--SIIPELEKAKSIQRP-----1302
Db      2353 SVNDHMSMAEAEIKRKIPALTESYONINNIVSYIAEAVKYLEIFDRDYGNOYQIVEHK 2412
QY      1302 --FSTFRIDNTGTGPPEALVERIEKTEVEGE-----RIIKEYEEELIEEVEREKV--GR 1353
Db      2413 KQSFILDRTN-----ALMDIEIFKKNNNNYNLWEVNETIHRYANDYEIKTNNVOAK 2466
QY      1354 TE-----WTLFQR---LAGFNELKODEENFYRELSEEM-----EFIDE 1390
Db      2467 TEYDIENIKQNDMDLQNIFLKFKSVILEYFENVKKRKSILNDLYEDERLLKIGEHDE 2526
QY      1391 IKMEASE 1397
Db      2527 IKNRYTE 2533

RESULT      3
TECU_HSVSA          STANDARD;           PRT; 2469 AA.
ID      TEGU_HSVSA
AC      Q01056        STANDARD;
DC      01-APR-1993 (REL. 25, CREATED)
DT      01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT      01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE      PROBABLE LARGE TEGUMENT PROTEIN.
DE      64 OR EEPF2.
OS      HERPESVIRUS SAIMIRI (STRAIN 11).
OC      VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC      GAMMAHERPESVIRINAE.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE: 92333688.
RA      ALBRECHT J.-C., NICHOLAS J., BILDER D., CAMERON K.R., BIESINGER B.,
RA      NEWMAN C., WITTMANN S., CRAXTON M.A., COLEMAN H., FLECKENSTEIN B.,
RA      HONESS R.W.;
RT      "Primary structure of the herpesvirus saimiri genome.";
RL      J. VIROL. 66:5047-5058(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE: 92230229.
RA      NICHOLAS J., CAMERON K.R., COLEMAN H., NEWMAN C., HONESS R.W.;
RT      "Analysis of nucleotide sequence of the rightmost 4.3 kbp of
RT      herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT      organization between HVS and Epstein-Barr virus.";
RL      VIROLOGY 188:296-310(1992).
CC      -1- FUNCTION: TEGUMENT PROTEIN.
CC      -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC      BHV-1 24, EBV BPFL1, HVS-1 64, VZV 22, AND HCMV UL48.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC      or send an email to license@isb-sib.ch).
CC      -----
Cc      EMBL: X64346; G67385;
DR      EMBL: M86409; G3:0993;
DR      PIR: H36812; H36812.
SQ      SEQUENCE 2469 AA; 280165 MW; 1B72F5F9 CRC32;

Query Match      2.1%; Score 155; DB 1; Length 2469;
Best Local Similarity 17.3%; Pred. No. 0.51;
Matches 230; Conservative 208; Mismatches 497; Indels 396; Gaps 58

QY      271 FTFSAGTVFVDPPDLGALYNHDEFHYODWTYLSQLENKPBDYIFYYYR-----324
Db      178 FTSNVNDAIEYLSPNNQYT--GSFLYFVKEKHSH-----YIMNHRYINTEKL 228
QY      324 -GNMDADGVGGSVLY-----TRSLTPESLIPLMQKAANSVGCRDFENNFI 367

```

```

Db 229 HGENIDLTSGEIIIEISPTNPKPTSTQPKPKPTPKATPKAPKPTPKPKESTI 288
QY 368 TTNSGCEPPLVERLEKTAEGEKLLEAV-----EIEVEKEVERKPT--EMTL 419
Db 289 PYKS--KRP--KIPKTSKSKKVLKRLDALPPOHKTEIEHRELLPTEFEVDNLT 343
QY 420 POR-----LLEGFKELO-----ODEENFVRELSEKEEILINLOMATEVEKL 462
Db 344 FHPVERITPTGDTLSLGINSTTKREDDEDDVNTSKLEDEDDWLD----- 393
QY 463 FGRALPIKRLRALAPHSNPLANHETIKYVYGSKLPGHKRFSWGMEDYFSIYVA----- 518
Db 393 ----IPIEVLDTETTHS-----DOETI-YMIGDE--NIMHMSYSDDDIDIDISTIQD 441
QY 518 -KICSSRIPIRYFRKSPRICGLDSRGLOLFSHGK-----NLSAHSINONVPGK 567
Db 442 NLITSLDNPKN--NTPRI--IDKTSNPIKEGKALHSIDRLKIVLEHGLITSSIS 497
QY 568 NSGCFPRPDVALMWEKKGOPAKTAIYAFILSVASKADAVALKTCGLKECEKRELEAK 627
Db 498 ISKCSLLOEVL--WGE-----KLSIPTR-DIKTIKTELITIEIAIALTK 542
QY 628 CISPACANAVACLOTCNNRPDETCQIKGDLFENSVDENECVSKKCVPRKSDVG 687
Db 543 -LTN-----DIFRNNAVITKLNKCMKLK-----SESYD 569
QY 688 DEP-----VDPDSVLOKFKM-----KDFS-----GKWFTRGLNPTF 720
Db 570 SYKHLASALNNITIKIOTIDTEIELKTLSTVTELGKDFSVVCKKSEITMAIKNKL 629
QY 721 DAFDCOLHEFTENKLVNLSWIRTPGCGFTRSAVKF-----VODPYRPG 769
Db 630 EKISTRKQELHTENYFQSVL-----IAMEPTOPIPLPRVIEIOPSKRAQ 675
QY 770 ILVHNDHEVILYODDWITLSSKVENSPEDYIEVYKGRND--AMDYGGSGVLTFRSAVL 826
Db 676 QL--HEKSKLVEOK--LTIANNVLTGLHTMKQDXTDISPAD-----FTYVLKN 722
QY 827 PESIIPLOTAQKVRD--FNTFIKIDNTGPPPLVERLEKKEVEGERTI--IKEVE 882
Db 723 IQSTHLOLCVTDNLIDKKFTISNTVOOLSTYGMVAELSHSONFPRADVPVPLKILDD 782
QY 883 IEBEVEKVDKEVTLFSKLFESFELORDENEFLELSKEMADVLDGKMEATEVEKLG 942
Db 783 IKKEICVTTKO--KNEETLSKILADVOLLLENASOSTLSIPLOIYITKAGTLVG 837
QY 943 RALPIKMAVAATHCFTSPCHDRIRFESSDDGIGRLGIT--RKRNGFLILKILPPIQ 998
Db 838 E-----RENOKFESLKNVTOKLSTSEFLKTLIDSTLENNVOLOIQ 878
QY 999 S-ADRLTGGSSRPLSAFRSGFSKG--IPDIPLPSKNELEKELAPLL----- 1046
Db 879 EISDILQSNQYHOSEITKQAFEDKSNITINNLOLQNNQYTYVOPMLIANKRLSRA 938
QY 1046 -----KLVGLACAFILVPSADAVDAKTCACILACKCRTELAKCIANPACANVAC 1096
Db 939 KPRESENTICIIITSLVSGLSLKSSTVEAKALKSIDLKEKL----- 984
QY 1097 LOTCNRPDETE-----COIKGDLFENSVDENECVSKKCVPRKSLGEPAPD 1149
Db 994 --TAVDRPLKRELYNVIRKLOKLTLEQOEEDNM-----KMEVDSF--VPT 1027
QY 1150 PSVAVONF--NISDFNGKWITSGLNTFPAPDCQHEHTEGDN-KLVGNISMRKILTD 1206
Db 1028 PSRDVKITFIONAPSKAKOYAKKALKDOIOAMEIDVDEPSEVEDINRANGQAMQ----- 1083
QY 1207 SGFTFSAVOKFVODPNQPGVLYNHDNEYLHYODDWIILSSKIEKKPEYIF--VYI 1261
Db 1083 -----KISAFODLNFSLI-----PDMWLSL-AKEYIRPKSLFTYVIGPILL 1124
QY 1262 RGRNDAMDGVGAIVYTRSSVLPNSIT-----PELEKAAK 1296
Db 1125 KFEVEVLESVKNTKEAKLSLJLNGPVFTPPKFDWLIHYEVSVNFLKTLINLPRVSVIAH 1184

```

```

QY 1297 SIGRDSSTIRFNDNT-CGEPALVERIE-----KIVE-----EGERILYKEV 1337
Db 1185 NICHSLSLSQLNLSKLTLEFAVGTSLBOHAKSCFMKLENTWHDHOVDTKTKIDEYI 1244
QY 1338 EEEIEEVER-----EVENVGR-----TENTLFOQLAEQFNELKODEENFVRELSEKEMEFL 1388
Db 1245 EDLRNTRKHIAVAPQIOSPNRFLSPEDIQELINSPLKLRDLSLENNESRLASQKNFQWL 1304
QY 1389 DELKMEASEVE 1399
Db 1305 -ENTVKAAELO 1314

RESULT 4
MYSB_PIG STANDARD; PRT; 1935 AA.
AC P79293;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH.
OS SUS SCROPA (PIG);
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
NC ARTIODACTYLA; SUIDFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DOMESTICA;
RA KO Y.L.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY
CC AND 1 HEAVY MEROMYOSIN (HMM) IT CAN LATER BE SPLIT FURTHER INTO
CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE MYOSIN, WHILE THE
CC BETA ISOFORM IS A 'SLOW' ATPASE.
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75316; G1698895;
DR PFAM; PF00063; myosin_head; 1.
DR KMW; PF00612; IQ; 1.
KW MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;
KW ATP-BINDING; METHYLATION; ALKYLATION; MOLLUSCEN FAMILLY.
FT DOMAIN 1 838
FT DOMAIN 1 838
FT DOMAIN 839 1935
FT DOMAIN 780 802
FT DOMAIN 839 1935
FT NP_BIND 178 185
FT DOMAIN 655 677
FT DOMAIN 757 777
FT MOD_RES 129 129
FT MOD_RES 695 695
FT MOD_RES 705 705
FT SEQUENCE 1935 AA; 223109 MW; F072F29E CRC32;

```


Query Match 2.0%; Score 148.5; DB 1; Length 1935;
 Best Local Similarity 16.7%; Pred. No. 0.87;
 Matches 248; Conservative 220; Mismatches 555; Indels 459; Gaps 57;

QY 166 PDTECOIKGDLFENSVDQNECAVSRKCCVPRKSDVGEFPYDRN-----AVY 216
 375 PDTE-----EADKSAVLMGINSADILKGLCHPRKVGNGENVTYKGVNQVQVATGALA 428
 DB 217 QANMADFSQKWITISGLNPT-----FDAFDCQLEH-----FHMENDKL 255
 DB 429 KAVEKMFN--WAVTR--INTLETKOPROYFTGVLDIAGFEIFEDNNSQOLINTNKL 485
 QY 256 -----VGNLTWIKTLDGGEFTSVAQTF-----V 280
 DB 486 QQFNNHMFVLEQBEYKKEGIEWEF--IDFGMDLOACIDILIEKPMGINSILEECMPKA 543
 QY 281 QDDPLGALY--NHDNEFLHVDQDWTLLSSQIENKPD-DYIFYYGRND-----AMDYGG 334
 DB 544 TDMTFKAKLDNHLGKSNFQK-----PRNKGPRPAHALIHAGTYDYNIGW----- 594
 QY 335 SVITRSPTLPESIIPILOKAA--KSYGRDFNNEITTDNSCGEPPLVERLEKTAEGEKL 393
 DB 594 --LQKNPDPLNEYVDLYKSSILKLSNLFANY-----AGADTP--VERKGKAKKGSF 644
 QY 394 LIKEAVEIEEVEKEVEKVDTEMTLFLQLLEGFKLODEENFVRELKKEKE----- 448
 DB 645 QTVSAH--RENLMKMTNLRST-----HFAFVACIIPNEIKSPGVID 685
 QY 448 ---ILNELQMEAT--EVEKLEGRALP-----IRKLRLMALPHSNFLANHETIKYY 492
 DB 666 NPLVNHQLRNGVLEGRICKGFPNRLILYGDPRQRYRLINPAIRPEGQIDSRKAEKL 745
 QY 493 VQSKLPGRKPSWGMEDYFGSIYVAKICSSRRIPRIYRKSPRICCGIDSGLOLFESHGH 552
 DB 746 LGSIDIDHNOYKFGHTKVFKAAGLLGLEEKDERLRLRIITRI--QAQSGVSRMEFK 803
 QY 553 NLSPAHS---INQVPRKNSCGKPPKQVALVWEMKGFATAVAIFLISVASKADAVD 609
 DB 804 LLERDSLLITQWNR-----AFMSVKNW--PWKLYIKIRPLESATEKEMA 850
 QY 610 ALKTCCLKEC-----RLELAKCISNPACANVACLOTNNRPDETCQIKGDLFE 662
 DB 851 TYKEEGRKLEKSEARKELEKMYSLQENKDLOLQVADQDMLISSEKCDOLIK 910
 QY 663 NSV-----VDEPNECAVSRK-----CYPRKSDVGEFPVPSVLYQKFD 702
 DB 911 NKIOLEAKYKEMTERLEDEEENNAELLAKKRVDECESELRDIDLELITAKYERK-- 969
 QY 703 MKDFSGKWTITGLNPTPAFCQLEHFEHTENKLVGNLSWRIRTPDGGFTSVAQV 762
 DB 969 -----HATELNK--VKNLTEEMGLD-----ELIAKLT 993
 QY 763 QDEKYPGILYNHDNEYLLYQDDWYLLSSKVENSPEDYIFVYKGNAMDYGGSVLYTR 822
 DB 994 KEKKAQEOAHOQALDLOAEDEKVNLTAKAYKLEOHV-----DLAAGLEOEK 1042
 QY 823 SA-----VLPESITPELOTAAQVQ-----RDF-----NTEFKIDNTGCP 857
 DB 1043 KYRMDLERAKRLEGLDKLTOESIMDLENDKOOLDERLKKKPFELNALNAREDEQALGS 1102
 QY 858 E-----PPLVERLEKKEVEGERT-----IKEYEVELEVEYK----- 890
 DB 1103 QLOKRLKELOARLEEEEL--EAERTARAKVERKLSRDLSELEESERLEEGAGATSVQI 1161
 QY 890 -----VRDEKVTLSKLPFGKRELQDEENFREL 919
 DB 1162 EMNKKREAEQKMRDLEATLOHEATAALRKKHADVAELGEQIDNOROKOLEKEX 1221
 QY 920 SKEEMVDVIGLMEATEVERLFGALPPIRKLMVATHCFTSPCHDIRFSSDDGIGRLG 979
 DB 1222 SEFKLEDD-----VTSNMEOIKAKANLEKX-----CRT--DEDMNHRS-----KAE 1264

QY 980 ITRKRINGTFLIKILPIOSADLRTTGGSSRPL--SAFRSGFSFGIDYVLPBKNEL 1036
 DB 1265 EYQSRVNDL-----ISQRAKLOTENGELSRLQDEKALISOLITRG-----KLFYTOOL 1312
 QY 1037 KELLAPLLIKLVYACAFLLVPSADAVDALKTACCLCKCRLEIAKCIANPACANVAC 1096
 DB 1313 EDLKRQLEEEYKAKNALAHALOSARHADLLREOYEETETKAELOVLSK--ANSEVAQ 1370
 QY 1097 LQTCNNRPETCOQKCGDLFENSVDNECAVSRKCCVPRKSDVGEFPAPPSVLYON 1156
 DB 1371 WRT-----KYETDALQRTLELEKAKKLAQRLQDADE----- 1403
 QY 1157 FNISDNGKWYITSGINPTFADFCQLEHFEHTEGDNKLVGNISKRITLDSGFTTSVAVQ 1216
 DB 1403 -----AVEAVNAKCSLEKTKHRLQNEIEDLWVD-----VERSNAA 1438
 QY 1217 KFYVDNQGCVLYNHDNEFLHVDQDWTLLSSKIEKRPDYIFYYGRNDAMDYGGAVY 1276
 DB 1439 AALDRKQ-----RNFPKILAEWKQYESSQLESQKE-----ARSL 1477
 QY 1277 YTRSSVLPNSIIPLE-----KAAKSIGRDFSTFRIDNTGCPPALVERIEKTEVEGE 1330
 DB 1478 STELFKAKNAVEESLEHLETSKREKNLOEETISDLEQDLSGSGKTILHELEKVRQLEAK 1537
 QY 1331 RIIVKEVEIEEVEKEVEKVGRTENTLQ-----RLAEGFNELKODEENFVRELS-- 1382
 DB 1538 LEIOSALEESEASLEHEEGKILPAQLEFNOIKAEERKLAEDMEQAKNNHLRYVDSL 1597
 QY 1382 -----KEEME--FLDEIKMEASEVEKLEFGKA 1405
 DB 1598 QTSIDATRSRNEALRYKKNMEGDNLNEMEIQLSHANMAAEA 1639

RESULT 5
 MYSB_HUMAN STANDARD; PRT; 1935 AA.
 AC P12883; Q14904; Q16579;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
 GN MYH7 OR MYHCB.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RN SEQUENCE FROM N.A.
 RN MEDLINE: 91065634.
 RA JAEINCKE T., DIEBERICH K.W., HAAS W., SCHLEICH J., LICHTER P.,
 RA PRODR T. M., BACH A., VOSBERG H.P.;
 RT "The complete sequence of the human beta-myosin heavy chain gene and
 RT a comparative analysis of its product.";
 RN GENOMICS 8:194-206(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RN MEDLINE: 90301496.
 RA LIEM C.-C., SOLE M.J., YAMAUCHI-TAKIHARA K., KELLAM B.,
 RA ANDERSON D.H., LIN L., LIEM J.;
 RT "Complete sequence and organization of the human cardiac beta-myosin
 RT heavy chain gene.";
 RN NUCLEIC ACIDS RES. 18:3647-3651(1990).
 RN [3]
 RN SEQUENCE OF 1-115 FROM N.A.
 RN MEDLINE: 89264452.
 RA YAMAUCHI-TAKIHARA K., SOLE M.J., LIEM J., ING D., LIEM C.-C.;
 RT "Characterization of human cardiac myosin heavy chain genes.";
 RN PROC. NATL. ACD. SCI. U.S.A. 86:3504-3508(1989).
 RN [4]
 RN ERRATUM.
 RA YAMAUCHI-TAKIHARA K., SOLE M.J., LIEM J., ING D., LIEM C.-C.;
 RN PROC. NATL. ACD. SCI. U.S.A. 86:7416-7417(1989).
 RN [5]
 RN SEQUENCE OF 1310-1935 FROM N.A.

RX MEDLINE; 86176778.
 RA SAEZ L., LEINWAND L.A.;
 RT "Characterization of diverse forms of myosin heavy chain expressed in
 RT adult human skeletal muscle.";
 RL NUCLEIC ACIDS RES. 14:2951-2965(1986).
 RN [6]
 RP REVISIONS.
 RA LEINWAND L.A.;
 RL SUBMITTED (MAR-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [7]
 RP SEQUENCE OF 1410-1935 FROM N.A.
 RX MEDLINE; 88299163.
 RA KURABAYASHI M., TSUCHIMOTO H., KOMURO I., TAKAKU F., YAZAKI Y.;
 RT "Molecular cloning and characterization of human cardiac alpha- and
 RT beta-form myosin heavy chain complementary DNA clones. Regulation of
 RT expression during development and pressure overload in human
 RT atrium.";
 RL J. CLIN. INVEST. 82:524-531(1988).
 RN [8]
 RP SEQUENCE OF 785-1935 FROM N.A.
 RX MEDLINE; 90235862.
 RA BOBER E., BUCHENGER-SEIDL A., BRAUN T., SINGH S., GOEDDE H.W.,
 RA ARNOLD H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL EUR. J. BIOCHEM. 189:55-65(1990).
 RN [9]
 RP SEQUENCE OF 1393-1935 FROM N.A.
 RX MEDLINE; 87192738.
 RA JANDRESKI M.A., LIEM C.-C.;
 RT "Construction of a human ventricular cDNA library and
 RT characterization of a beta myosin heavy chain cDNA clone.";
 RL HUM. GENET. 76:47-53(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE; 96039076.
 RA ARI S., MATSUOKA R., HIRAYAMA K., SUKURAI H., TAMURA M., OZAWA T.,
 RA KIMURA M., IMAMURA S.-I., FURUTANI Y., JOH-O K., KAWANA M., TAKAO A.,
 RA HOSODA S., NOMMA K.;
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
 RT hypertrophic cardiomyopathy.";
 RL AM. J. MED. GENET. 58:267-276(1995).
 RN [11]
 RP VARIANTS HCM GLU-256 AND ARG-741.
 RX MEDLINE; 93248216.
 RA FANANAPAZIR L., DALAKAS M.C., CYRAN F., COHN G., EPSTEIN N.D.;
 RT "Missense mutations in the beta-myosin heavy-chain gene cause central
 RT core disease in hypertrophic cardiomyopathy.";
 RL PROC. NATL. ACADE. SCI. U.S.A. 90:3993-3997(1993).
 RN [12]
 RP VARIANTS HCM GLN-403.
 RX MEDLINE; 90367131.
 RA GEISLERER-LORANCE A.A.T., KASS S., TANIGAWA G., VOSBERG H.-P.,
 RA MCKENNA W., SEIDMAN C.E., SEIDMAN J.G.;
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta
 RT cardiac myosin heavy chain gene missense mutation.";
 RL CELL 62:999-1006(1990).
 RN [13]
 RP VARIANTS HCM GLN-249; GLN-403; CYS-453 AND MET-606.
 RX MEDLINE; 92204193.
 RA WATKINS H., ROSENZWEIG A., HWANG D.S., LEVI T., MCKENNA W.,
 RA SEIDMAN C.E., SEIDMAN J.G.;
 RT "Characteristics and prognostic implications of myosin missense
 RT mutations in familial hypertrophic cardiomyopathy.";
 RL NEW ENGL. J. MED. 326:1108-1114(1992).
 RN [14]
 RP VARIANTS HCM GLN-403; CYS-453; ARG-584 AND MET-606.
 RX MEDLINE; 94070863.
 RA WATKINS H., THIERFELDER L., ANAN R., JARCHO J., MATSUMORI A.,
 RA MCKENNA W., SEIDMAN J.G., SEIDMAN C.E.;
 RT "Independent origin of identical beta cardiac myosin heavy-chain
 RT mutations in hypertrophic cardiomyopathy.";
 RL AM. J. HUM. GENET. 53:1180-1185(1993).
 RN [15]
 RP VARIANTS HCM GLN-403 AND VAL-908.
 RX MEDLINE; 92346810.
 RA EPSTEIN N.D., COHN G.M., CYRAN F., FANANAPAZIR L.;
 RT "Differences in clinical expression of hypertrophic cardiomyopathy
 RT associated with two distinct mutations in the beta-myosin heavy chain
 RT gene. A 908Leu->Val mutation and a 403Arg->Gln mutation.";
 RL CIRCULATION 86:345-352(1992).
 RN [16]
 RP VARIANTS HCM LEU-403 AND TRP-403.
 RX MEDLINE; 94075629.
 RA DAVISE E., KOMADA M., FETLER L., DUBOURG O., DUFOUR C., CARRIER L.,
 RA WISEMSKY C., BERCOVICI J., HENGSTENBERG C., AL-MADANI S.;
 RT "Familial hypertrophic cardiomyopathy: Microsatellite haplotyping and
 RT identification of a hot spot for mutations in the beta-myosin heavy
 RT chain gene.";
 RL J. CLIN. INVEST. 92:2807-2813(1993).
 RN [17]
 RP VARIANTS HCM TRP-403.
 RX MEDLINE; 94093568.
 RA MOOLMAN J.C., BRINK P.A., CORFIELD V.A.;
 RT "Identification of a new missense mutation at Arg403, a CpG mutation
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in
 RT hypertrophic cardiomyopathy.";
 RL HUM. MOL. GENET. 2:1731-1732(1993).
 RN [18]
 RP VARIANTS HCM ASN-615.
 RX MEDLINE; 93038688.
 RA NISHI H., KIMURA A., HARADA H., TOSHIMA H., SASAZUKI T.;
 RT "Novel missense mutation in cardiac beta-myosin heavy chain gene
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 188:379-387(1992).
 RN [19]
 RP VARIANTS HCM GLY-778.
 RX MEDLINE; 93343938.
 RA HARADA H., KIMURA A., NISHI H., SASAZUKI T., TOSHIMA H.;
 RT "A missense mutation of cardiac beta-myosin heavy chain gene linked
 RT to familial hypertrophic cardiomyopathy in affected Japanese
 RT families.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 194:791-798(1993).
 RN [20]
 RP VARIANTS HCM VAL-908.
 RX MEDLINE; 93168485.
 RA AL-MADANI S., CHAMBERLAIN S., CLELAND J., NIHOYANPOULOS P.,
 RA GILLIGAN D., FRENCH J., CHOUDHURY L., WILLIAMSON R., OAKLEY C.;
 RT "Identification of a mutation in the beta cardiac myosin heavy chain
 RT gene in a family with hypertrophic cardiomyopathy.";
 RL BR. HEART J. 69:136-141(1993).
 RN [21]
 RP VARIANTS HCM TRP-719.
 RX MEDLINE; 95179132.
 RA GREVE G., BACHINSKI L., FRIEDMAN D.L., CZERNIOWICZ G., ANAN R.,
 RA TOWBIN J., SEIDMAN C.E., ROBERTS R.;
 RT "Isolation of a de novo mutant myocardial beta MHC protein in a
 RT pedigree with hypertrophic cardiomyopathy.";
 RL HUM. MOL. GENET. 3:2073-2075(1994).
 RN [22]
 RP VARIANTS HCM CYS-513; ARG-716, AND TRP-719.
 RX MEDLINE; 94110336.
 RA ANAN R., GREVE G., THIERFELDER L., WATKINS H., MCKENNA W., SOLOMON S.,
 RA VECCHIO C., SHONO H., NAKAO S., TANAKA H., MARES A., JR., TOWBIN J.A.,
 RA SPIRITO P., ROBERTS R., SEIDMAN J.G., SEIDMAN C.E.;
 RT "Prognostic implications of novel beta cardiac myosin heavy chain gene
 RT mutations that cause familial hypertrophic cardiomyopathy.";
 RL J. CLIN. INVEST. 93:280-285(1994).
 RN [23]
 RP VARIANTS HCM THR-797.
 RX MEDLINE; 96047159.
 RA MOOLMAN J.C., BRINK P.A., CORFIELD V.A.;
 RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
 RT myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL HUM. MUTAT. 6:197-198(1995).

[24]
RN VARIANT HCM CYS-453.
RX MEDLINE: 96209901.
RA KO Y.-L., CHEN J.-J., TANG T.-K., CHENG J.-J., LIN S.-Y., LIU Y.-C.,
RA KUN P., WU C.-W., LIEN W.-P., LIEN C.-C.,
RA "Malignant familial hypertrophic cardiomyopathy in a family with a
RT 453Arg-->Cys mutation in the beta-myosin heavy chain gene:
RT coexistence of sudden death and end-stage heart failure."
RL HUM. GENET. 97:585-590(1996).
RN [25]
RX VARIANTS HCM THR-349 AND TRP-719.
RX MEDLINE: 98204402.
RA JESCHKE B., UHL K., WEIST B., SCHRODER D., METTINGER T.,
RA DOHLEMAN C., VOSBERG H.-P.,
RT "A high risk phenotype of hypertrophic cardiomyopathy associated with
RT a compound genotype of two mutated beta-myosin heavy chain genes."
RT

Query Match 1.9% Score 146; DB 1; Length 1935;
Best Local Similarity 16.9%; Pred. No. 1.2;
Matches 238; Conservative 207; Mismatches 529; Indels 434; Gaps 51;
QY 203 DYGEPPVDPNNAVQ-----NFNMDFSGKWTITSGLPNTPDAPDCQHEF 248
DB 461 DIAGEFIDPNSFEQCIATFNELQOFFNHHMEVLEQEEYKKGIETWIFIDFGMDLQAC 520
QY 249 HNEVDKLVG-----NLTKIKTLDDGFFTRSAVQTFVQDPLPGALYNHD 293
DB 521 IDLIEKPMGIMSLIEECMFPRATMTKAKLFD-----NHL 557
QY 294 NEFLHODDMYILISQIENKPD-DYIFVYGRND---AMDYGGSVIYTRSPLEPESI 348
DB 558 GKANQK-----PRIKKGPFAHSLIHAGIVDYNIIGW-----LQKNKPLNLETY 605
QY 349 IPNMQAA-KSVGRDFNNFTTNSCGPPPLVERLEKTAEGEKLKKEVEIEEVEK 407
DB 606 VGLYQKSSLELSTLFANY-----AGADAP-IEKGKGAKKSSFOVYSAH-RENLNK 657
QY 408 EYEVKVDTEMTLQRLLEGKELQODEENFEVLEKEKE-----LNELOMEAT- 458
DB 658 LMTNLKST-----HPIFVACIIPNETKSPGVNDPLVMQLNCVIL 699
QY 458 EVELKFGRLP-----IRKLMLALPHSNFLANHEITIKYVSKLPGHRRFSWG 506
DB 700 EGIKICKGPNNILYGDFFQRYRIINPAIPGQFIDSRKAEXKLSLIDHNOYKFG 759
QY 507 WEDYFGSVIAKICSSRIIRYFRKSPRICGSDSGQLDFSHGKHLNPAHS---INON 563
DB 760 HTKVFRAAGLGLLEEMRDELSRIITRI--QAOSRGVLAEMKYLERRDSLVLIQMN 817
QY 564 VPKNSGCKFPKDVALLVMEKMGOFAKTALVAFILSVASKDAVDALKTCTCLKEC-- 622
DB 818 IR-----AFMGVKNW-PMMKLYEKIKPLKSAERKEMASKEEFTRLKEALE 864
QY 622 -----RLFLAKCISNPACANVACLOTCNNRPETECQIKGGLFENSV---VDEENE 671
DB 865 KSEARKLEKEMVSLLOEKNDLOLOVQADQDNLADHEECODLKNKIQLEAKYKEMNE 924
QY 672 -----CAVSRK---KCVPRKSDVGFPPVPSVLYOKFDMKDFSGKMFITRGL 716
DB 935 RLEDEENNAELTAKKRLEDECELSKRDIDDELTLAKYKERE----- 969
QY 717 NPTDAPDCQHEFTTEENKLVNLSMRIRTPDGGFFTRSAVQTFVQDPLPGALYNHD 776
DB 969 -----HATENK-VKNILTEMGILD-----ELIKLTKERKALQEAHQOAL 1007
QY 777 EYLLYODDMYILISKVENSPEDYIFVYKGRNDAMDYGGSVIYTRSA----- 825
DB 1008 DDLQAEEDKVNLTAKAKKLEQV-----DLEGLLEDEKRYRMDLERAKKLE 1056
QY 825 ---VPESTIPELQTRAQVQ---RDV-----NFTIKDNTGPE-----PLIVE 863
DB 1057 GDLKLTQESIMLENDKQDLBERLKKKQFELNALMARIEDQALGSOLOKKEIQARIE 1116

QY 864 RLEKVEEGERT-----IIEVEIEEVEK----- 890
DB 1117 ELEBEL-ESERIRANAVENKRLSDLSRELEISERLEEAAGATSVQIEMKKREAEFOKMR 1175
QY 890 -----VROKEVLFESKLEFGRELDORDEENLRELKSEMDVLGLKME 933
DB 1176 RDLEATLOHEATAAALKHKADSVALEQIDNLORVKOKLEKESKLELDD-----V 1231
QY 934 ATEVEKLEGRALPRKLMVAATHCFSPCHDIRFSSDDGIGRLGTRKRINGTFLUKI 993
DB 1232 TSNEQIILKAKANLEKMA-----CRT-LEDQNNHRS-----KAEIQRVNDL----- 1274
QY 994 LPPIQSADLRITVGRSSRP---SAFRSGFGKIFDIIVPLPSKNEKELTAPLLIKVY 1050
DB 1274 -TSQRAKLOTENGLSELROLDKEKALISQLTRG-----KITTYQLEDLKROLEEVKAK 1326
QY 1051 LACAFILYPSAIVADALKTACALLKGRIFLACIANPACANVACLOTCNNRPDETECO 1110
DB 1327 NALAHALQSAHDDDLREQYEETEKAKELORYLSK--ANSEVAQMR----- 1374
QY 1111 IKCGDLFENSVYDEFNECAVSRRKKCVPRKSDLGEFPAPDPVLYQNFNSDNGKWTITS 1170
DB 1374 -----KYETDAIQRTLEELFEAKKTLAQRLQEAEE-----AVE 1405
QY 1171 GLNPTFADPCQHEFTTEGNKLVGNISMRITKLDGFFTRSAVQTFVQDPLPGALYN 1230
DB 1406 AVNACSSLEKTRHRLQNEIEDLVND-----VERSNAAAALDKQ-----RN 1448
QY 1231 HONEFLHODDMYILISQIENKPDYIFVYGRNDAMDYGGSVIYTRSPLEPESI 1290
DB 1449 FDKILAEKKCKYEESSQLESQKE-----ANSLTELFKLKNAYEES 1491
QY 1291 LE-----KAAKSTIGRDFSTFRIDNTGCPPEPALVERLEKTEVEGEITIKYVEIEEVE 1344
DB 1492 LEHLETFKRENKLNLOEISDLTEQLGSSGKTIHLEKVRQKLEKEMLOSALBEAEASL 1551
QY 1345 EKEVEVGTETLQ-----RLAEGFNELODEENFEVLEK----- 1382
DB 1552 EHEEGKILAQLEFNQIKAEIRKLAEKDEMEQAKRHHLRVNDSLOTSDAETRSRNEA 1611
QY 1382 ---KEEME-FIDIEKMEASEVEKLEFGKA 1405
DB 1612 LRVKMKMEDLDMEMELQLSHANMAAEA 1639
RESULT 6
ID NUM1 YEAST STANDARD; PRT: 2748 AA.
AC 000402;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE NUCLEAR MIGRATION PROTEIN NUM1.
GN NUM1 OR YDR150W.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EIKAROTR: FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 28383 / FL100;
RX MEDLINE: 92079907.
RA KOMANEK J., SCHAAF-GERSTENSCHLAGER I., ZIMMERMANN F.K.,
RA PERECKO D., KUENZEL H.;
RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the
RT highly repetitive 313 kDa NUM1 protein."
RL HUM. GEN. 230:277-287(1991).
CC -I- FUNCTION: CONTROLS NUCLEAR MIGRATION. NUM1 SPECIFICALLY CONTROLS
CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
CC CYTOPLASMIC X-CROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
CC ENVELOPE.
CC -I- ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT CONSENSUS
CC (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH FLANKING

DOMAINS OF THE TANDEM REPEATS.
 CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X6136; G4072; .
 CC PIR: S19052; S19052.
 CC SCD: L0001287; NM1.
 CC PROSITE: PS50003; PH_DOMAIN: 1.
 CC PFAM: PF00169; PH: 1.
 KW REPEAT.
 FT DOMAIN 593 1384 12.5 X TANDEM REPEATS.
 FT REPEAT 593 656 1.
 FT REPEAT 657 727 2.
 FT REPEAT 728 798 3.
 FT REPEAT 799 862 4.
 FT REPEAT 863 925 5.
 FT REPEAT 927 990 6.
 FT REPEAT 991 1034 7.
 FT REPEAT 1035 1118 8.
 FT REPEAT 1119 1182 9.
 FT REPEAT 1183 1246 10.
 FT REPEAT 1247 1310 11.
 FT REPEAT 1311 1374 12.
 FT REPEAT 1375 1384 13 (INCOMPLETE).
 FT DOMAIN 2573 2683 PH.
 SO SEQUENCE 2748 AA; 313202 MW; 3085662C CRC32;

Query Match 1.9%; Score 143; DB 1; Length 2748;
 Best Local Similarity 17.9%; Pred No. 2.9; Mismatches 582; Indels 486; Gaps 72;
 Matches 283; Conservative 231;

QY 62 SSSSDSHCKDKSQCISIDTSEEFIOFPLKGMFLLEKOWRF-----IOLAIYLC 115
 DB 748 SAVEDLVCKENPDV-----EFLKESAKLGHITVASEESELQRYKSYLEKEVOPS 800
 QY 116 TFIYVRVAVNALKTCACCLKECRLELAKCIANPSC-----AANVACLOTNNRPDET 169
 DB 801 LAVIVHAKATDH-----HLSDSAVEELVKCKENPDMEFLKESAKLGHITVASEAYSEL 856
 QY 170 ECOL-----KCGD--LFENSVDQFNECAVSRKKCVPRKSPVGEFVPDRNAV 215
 DB 857 EKKLEOPSLAVLVERHAKATDHLLSDSAVEDLVCK-----KENDV--EF----- 900
 QY 216 YONFNMKDPSGKWIYITSGINPTFDALFDCQLHEFHENDKLVGNITWRKITLDGGFTRSA 275
 DB 900 -----LKESSAKLGHITVASEAYSELEKLEQ-----PSLAVLVERHAKATDHLLSDSA 948
 QY 276 VOTVODDPLGALYNHNEELHYO-----DDWYITSSQINRKDDYIFVYVYRG 324
 DB 949 YE-----DLVCKENPDMEFLKESAKLGHITVASEAYSELEKLEQPSLEVLVHAKA 1002
 QY 325 RND--AMDGGSV-----IYRSPITPESTIPN-----LOKAASVGRDF--- 364
 DB 1003 TNHLLSSAYEDLVCKENPDMEFLKESAKLGHITVASEAYSELEKLEQPSLEYLVE 1062
 QY 364 -----NNEITDNS-----CGPEPPIVERLEKTAEGEKILLIKEN--ELEEEVEE--- 409
 DB 1063 HAKATNHLLSDSAVEELVKCKENPDVLEKESAKLGHITVASEAYSELEKLEQPSLE 1122
 QY 409 --VEVVRTEMTLFLQLEGEKELQODEENFVRELSEKEEILNE--LOMEA--TEVERLEF 463
 DB 1123 YLVEAKATNHLLSD--SAVEELVKCKENPDVLEKESAKLGHITVASEAYSELEK--- 1179
 QY 464 GRALPIRKALMALAPHSNPLANHETIKYVYGSKLPGHKRISWGMEDYFSGIYVAK----- 519

DB 1179 -----KLEQ--PSLAVLVEH-----AAADHLLS--DSAYEDLVCKENPDV 1217
 QY 519 -----ICSSRRIPRYFRKSPRICGLDSRG--QLFSHGKHLSPASINONV 564
 DB 1218 EFLKESAKLGHITVASEAYSELEK-----LEQPSLAVLVERHAK--ADHLLSSA 1268
 QY 565 PKNSGCKFPKDYALVWMEKQOFKATLVAFLILSVASKAD-----AVDAKTCQCL 617
 DB 1269 YEDLVCKENPDMEFLK--EKSALGHITVASEAYSELEKLEQPSLEYLVEHAKATNHLL 1327
 QY 618 LKECRLE--LAKCISNPAC-----AANVACLOTNNRPDETFCQDKCDLFDNSVDFEN 670
 DB 1328 LSDSAVEDLVCKENPDMEFLKESAKLGHITVASEAYSELEKLEQPSL----- 1378
 QY 671 ECAVSRKKCVPRK--SVGPF-----PVDPVSVLYOKTDMKDFSGKWEITRGLNPTFPAF 723
 DB 1378 EYLVKHAEOIOSKIISISIPSTLANPSMEDMASKLQKLEYOIVSNDEYI--ALKNTWEKP 1435
 QY 724 DCQ-----LHEFH--TEENKLVGNLSWRIRTPDGGFTTSAYOKVPOD--PKYQGLYN 773
 DB 1436 DVELLSKTLKGYHIIDTTYNELVSNFN-----SPTLKEIEKAKSKGYRLI 1482
 QY 774 HDNEXL-----LYODWYILSSK-----YENSPE-----D 798
 DB 1483 EPREYIDLNRITATPSKEIDNFCQIOCVALDSKEYERLKNLSNPSKFEIENALLD 1542
 QY 799 YIFV--YKGRNDAMDVGSGSVLYTRSAVLPESITPELOTAQAQKVRDNTFIKIDNTC 855
 DB 1543 LVLYDTEYQAMKD-----NASKKSLIPSTKVL-----DVTM----- 1577
 QY 856 GPEPPIVERLEKVEGERTIIEVEEIEEVEKVRDREVTLSKLEFGFKELORD--- 912
 DB 1577 -PAPOLASAKESLSLOK-----RTLSDIENELKAL-----GVVALRKENLNP 1616
 QY 912 -EENFLRELSKEEM-----DIVDGLKMEATEVEKLFG-----RALPIR 948
 DB 1617 LKPIYDANAKNDVNLCSKSLVPLSTEYDNKRKHITLILDDPSIDFLKEXCEXY 1676
 QY 949 KLMAYATH-----CFTSPCHDIRREFSSDDGIGRLGIRKRIKNGTFLFKILPIQOSAD 1001
 DB 1677 QMLIISKDYEEKOEALENPGEYFELKASALGY-----ELVSEVLDKRMQIMDSDP 1729
 QY 1002 IRTTGRSSRPLASRSGFSKIGIDIYPLPSKNEIKELIAPLLKLVGLACAFILVPSA 1061
 DB 1730 IDYOKERARN-----EMVLL--RNEKE--ALOKKIEYPSLFTLEKAA 1770
 QY 1062 -----DAVDALTCACLLKGRLELAKCIANPACAAVACLOTNNRPDETEQOIKG 1114
 DB 1771 GNNKILVDCIEYDET-----IRK-----NPTMELEBSCH 1802
 QY 1115 DLFENSVDNECAVSRKKCVPRK--SDLGEFAPDPVSVLVONFINSDFNGKWIYITSGIN 1173
 DB 1803 HL--NLVLDNNEYSTIREPLKNNVEDLINTLSKLYAIPNTIYODIDIGXE-----N 1855
 QY 1174 PTDFAPDCQLHEFHTEGDNKLVGNISWRITKTLDSGFTTSAYOKVPODNOGCVLY--- 1230
 DB 1856 PNFYDLKDSLNK-----MDYVALISQDVELMAYATEKQOLDYLIKIS 1896
 QY 1230 -----NHDEYIHYODDWYILSSKIENKPEYIFVYVYRGANDAMDG 1270
 DB 1897 SEKIDHIVLPSYNNLMVTNRYNPSLSYLKEKAVLNHLLIKEDDYKNIL-----AVSE 1990
 QY 1271 YGGAVVYTRSSVLPNSILPELEKAASIGRDSFTIRIDNTCGPPALVERLEKTEVEGE 1330
 DB 1951 HPYVYIHLSERHLLNKNVVD-----ANDYATYRSKIE--KPTIDELSTALSG- 1998
 QY 1331 RIIVKEY-----EIEEVEKEVEKVGRTMTLFLORLAGFENELKODEENFVRELSEKEM 1385
 DB 1998 KILVNSTEHRNKKLSEPDSEFLMKAKEOGLIISKEYSELRDQIDRNDNDVLAKEA 2057
 QY 1386 EFLDEIKMEASEVEKLFQKALP 1407

```

RESULT 7
ID MYSB_RAT STANDARD: PRT: 1935 AA.
AC P02564;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH7.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE: 90016823.
RA KRAAT R., BRAVO-ZEHNDER M., TAYLOR D., LEINWAND L.A.;
RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences. Complete nucleotide sequence of full length cDNA for rat beta cardiac myosin heavy chain."
RL NUCLEIC ACIDS RES. 17:7529-7530(1989).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE: 90133919.
RA MCNALLY E.M., KRAAT R., BRAVO-ZEHNDER M., TAYLOR D., LEINWAND L.A.;
RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparisons suggest a molecular basis for functional differences."
RL J. MOL. BIOL. 210:665-671(1988).
RN [3]
RP SEQUENCE OF 1524-1935 FROM N.A.
RX MEDLINE: 82220036.
RA MAHDAYI V., PERISANY M., NADAL-GINARD B.;
RT "Molecular characterization of two myosin heavy chain genes expressed in the adult heart."
RL NATURE 297:659-664(1982).
RN [4]
RP SEQUENCE OF 1871-1935 FROM N.A.
RX STRAIN=WISTAR; TISSUE=HEART;
RC MEDLINE: 85179510.
RA MAHDAYI V., LOMPRE A.M., CHAMBERS A.P., NADAL-GINARD B.;
RT "Cardiac myosin heavy chain isozymic transitions during development and under pathological conditions are regulated at the level of mRNA availability."
RL EUR. HEART J. 5:181-191(1984).
RN [5]
RP EUR. HEART J. 5:181-191(1984).
RT FUNCTION: MUSCLE CONTRACTION.
CC -1 SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1 SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1 DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES.
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1 PM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1 EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1 THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
CC -1 THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN THE RAT. TWO OF WHICH ARE SPECIFIC FOR ADULT CARDIAC MYOSIN HEAVY CHAINS.
CC -1 SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES, WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE CONSERVED.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

CC	or send an email to license@stdb.cn).
DR	EMBL: X15939; G56657; -
DR	EMBL: J00752; G205578; -
DR	EMBL: M32698; G205599; -
DR	PIR: S06006; S06006.
DR	PIR: A02989; A02989.
DR	PFAM: PF00063; myosin_head; 1.
DR	PFAM: PF00612; 10; 1.
DR	HSPB: P08799; 10N1.
KM	MOSSIN: MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING; ALKYLATION; MULTIGENE FAMILY.
KM	ALP-BINDING; METHYLATION; ALKYLATION; MULTIGENE FAMILY.
FT	DOMAIN 1 839
FT	DOMAIN 840 1935
FT	DOMAIN 780 802
FT	DOMAIN 840 1935
FT	NP_BIND 178 185
FT	DOMAIN 655 677
FT	MOD_RES 129 771
FT	MOD_RES 695 129
FT	MOD_RES 757 705
FT	MOD_RES 705 1531
FT	CONFLICT 1529 1531
FT	CONFLICT 1731 1731
FT	CONFLICT 1784 1784
FT	CONFLICT 1851 1851
FT	CONFLICT 1858 1858
SEQ	SEQUENCE 1935 AA: 223082 MW: 1D16F5E CRC32:

[illegible]

```

QY 663 NSV-----VDEFNE-----CAVSRC---KQVPRKSDVGDPEVPDPVYLQKFD 702
DB 911 NKIOLEAVKEMTERLEDEDEEMAEITAKRRKLEDECSSEKRDIDLEITLAKVEKER-- 969
QY 703 MKDSGKWFITRGINTFADFCQLEHFEHTEENKLVGNLSWIRTRIPDGGFTFRSAVQKV 762
DB 969 -----HATENK--KNTITEMAGID-----ETIYKLT 993
QY 763 QDPKPGILYNHNEVYLLODDWYIISKVENSPEDYIFVYVYKGRNDAMDYGGSVLYTR 822
DB 994 KKKALDEAHQOALDLOABEDVNTLTAKVKYLEQVY-----DDLESLODK 1042
QY 823 SA-----VLPESILPELOTAACKV-----RDE-----NTIKDNTICP 857
DB 1043 KVRMDLERAKRLEGLDKLTQESIMLENDKQLODERLKKQDFELMAMINARIEDBOALGS 1102
QY 858 E-----PIVERLEKKVEGERT-----IKEVEELEEEVER----- 890
DB 1103 QLOKKIKLEOARIELELEEL-BAERTARAKVEKLRSDLSRELFEISRLIEAGGATSVQI 1161
QY 890 -----VRKEVTLFSKLFEGFELQDEENFLREI 919
DB 1162 ENKKRREAFOKMRDLEATLOHEATAALRKHAHSVABLGQIDNIQVRVKLEKER 1221
QY 920 SKEMOVLDGLKMEATEVEKLEFGRLPIRKLMVAVHCTSPCHDRIFRSSDDGIGRIG 979
DB 1222 SEFKLELD-----VTSNNEQIITAKANLEK-----CRT--LEDDQNHRS-----KAE 1264
QY 980 ITRKRNGTFELIKILPTQSDALRTTGGRSSRP--SAPRSFGSGIIDIYVLSKNEI 1036
DB 1265 ETQSRVNDL-----TRQRAKLOTENGELSRQDEKALISQING--KLITYTOOL 1312
QY 1037 KELLAPFLIKLVYLACAFLLVPSADAVDALKTACILKGCRIELAKCIANPACANVAC 1096
DB 1313 EDLRQLEEEVKAKNALAHALOSARHDCDLRQYEETEAKEALORVLSK--ANSEVAQ 1370
QY 1097 LOTCNRPDETECOIKCGDLEFNSVDFNECAVSRKCVPRSDGEPADPSVLYQN 1156
DB 1371 WRT-----KYETDALQRTLEELAEAKKIAORLODAE----- 1403
QY 1157 FNISDFNCKWYITSGLNPTDAFDQCHHEFHTEGDNKLVGNISWRIKTLDSGFTRSVQ 1216
DB 1403 -----AVEAVAKCSSLEKTRHRLQNEIEDLMDV-----VERSMAA 1438
QY 1217 KFDVDPNQPQVLYNHNDVLYHODWYIISKLENKPEDYIFVYVYKGRNDAMDYGGAIV 1276
DB 1439 AALADKRO-----RNFEDIIVEMKQKYEESQSELESQKE-----ARSL 1477
QY 1277 YTRSSVLPNSIIPLE-----KAAKSIGRDPSTFRIDNTCGEPALVERIEKTEVEGE 1330
DB 1478 STEFLKAKNAVEESLEHLETERENKNLOEISDLTEQIGSTGKSHELEKIRQOLEAK 1337
QY 1331 RIIVKEVEELEEEVEKEVEKVGRTENTLFQ-----RLAEGFELQODEENFYREL-- 1382
DB 1538 LELQSLAEAEASLEHEEKILIRAOLEFNQIKATELEKLADEMOARNMHLRYVDSL 1597
QY 1382 -----XEME-FUDELTKMASEVEKLFCKA 1405
DB 1598 QTSLOAETSRNEALVRKKMKMGDLNEMEIOLSHANMAA 1639

```

```

OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RA MEDLINE; 93247549.
RT KOELLING R., NGUYEN T., CHEN E.Y., BOTSTEIN D.;
RL "A new yeast gene with a myosin-like heptad repeat structure.";
RN MOL. GEN. GENET. 237:359-369(1993).
RP SEQUENCE FROM N.A.
RA MEDLINE; 94205265.
RA BOU G., ESTEBAN P.F., BALADRON V., GONZALEZ G.A., CANTALEJO J.G.,
RA REMACHA M., JIMENEZ A., DEL REY F., BALLESTA J.P.G., REVUELTA J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RL new open reading frames.";
RL YAST 9:1349-1354(1993).
CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -1- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; L01992; G171959; -
DR EMBL; X73541; G450554; -
DR EMBL; Z28320; G485587; -
DR PIR; S38173; S38173.
DR SGD; L0001122; MPL1.
KW MYOSIN; HEPTAD REPEAT PATTERN; COILED COIL; DNA REPAIR.
FT DOMAIN 69 487
FT 531 1678 COILED COIL (POTENTIAL).
FT 1834 1866 COILED COIL (POTENTIAL).
FT 301 301 R -> A (IN REF. 1).
FT 1875 AA; 21845 MW; 8801FDDO CRC32;
SQ SEQUENCE

Query Match 1.8%; Score 137.5; DB 1; Length 1875;
Best Local Similarity 17.5%; Pred. No. 3.6;
Matches 252; Conservative 240; Mismatches 525; Indels 421; Gaps 67;

QY 171 CQIKGDLFENSVD-----QFNEC-----AVSRKKCVPRKSDVGEFP 208
DB 30 CSLEOVKSEFGDVYKRLNDKLLQFNEKLSNMLKTVSFDLKAASSLKIDGLKTEM-ENV 88
QY 209 VPDNRNAV-----VONFMKDPFSGKWTITGLN-----FFDAF 241
DB 89 IRENDKIRKENDTVYKESVENEKMLSSLEFVKRKLDDLFEKKETQSNQORLKL 148
QY 242 DCQLHEFH-----KLVGNLTWRIKTLDDGEF--ASAVQTYQDDPLG 287
DB 149 DERLKEILELVRENNRSNSECKLKRLSTIMOLETYQOGYITINDINSRLEKRTQELTL-- 207
QY 288 ALYNHNEFLHODWYIILSQIENKRPDDYIFVYVYKGRNDAMDYGGSVLYTRSPILPES 347
DB 207 -----LQSNNDW--LEKELRSKNQOYL--SYROKTD-----K 234
QY 348 IIPNLQAKASVGRDF-----NFIITDNSCGPEPLVERLEKTAEGEKLKAVEITE 402
DB 235 VILDIRNELNRLRDLQFQERNNNDVLYKQK-----NELSKSLQ--EKL--BIKGLS 282
QY 403 EEVEKEVAVDTMTLFOR-----LBSGFKLODEENFYREL-----SKREK 446
DB 283 DSLNSEKQEF--SAMSLSKQRLVLDLSQLVNAVKEELNSIRELNTAKVLAJDSKKQTPENE 341
QY 447 EILNLOEATEVEKLEGRALPIRKLR--MALAPHSFLANHETIYYGSKLPGRKRF 505

```

```

Db 342 DLKLELOITREKLAQCEKCLRLSTITDEADENENLSAKSSDFIFLKQIKERITE 401
QY 506 GWEDYFGSIVAKICSSRRIPRIYFRKSPRICGDSGLQFSGHGHNSPHASINOWVP 565
Db 402 HLONOIEFIVELEHAKYPIINSFKERTMLENELNNALEHTSEKAKKKEIN---A 458
QY 566 KNSGCKEPPKDYALMVMWKKW-----GOFKTAIYAFILVASKA 605
Db 459 KNOCLVECBNDQITKRDLCROIQYLLITNSVSDSGPRKEEIO--FIONIMED 516
QY 606 DAV---DALKTCTCLKECR--ELAKCISNPACANVACLOTNNRPDEPCQIKG- 659
Db 517 DSITTESDQKVTLEELVEFKNIIOLOE-----KNALLKVRNLADKLESKEKSK 568
QY 659 ---DLFENSVDPEFNCASRRKCVPRKSDYD-----FPVPDPVLVQKED 702
Db 569 OSLOKIESETVNEAKEAITLKS---EKMDLESRIEQLKELEKTSVNEASYSNT 625
QY 703 MKDFSGKWFITGLNTPEDAFDQLEHTEENKLVNLSKRITPDGCGFTSNAVQEV 762
Db 626 IKQ-----LLETKRDLSEVOQLQTRISQIT-----RESTEN---MSLINKET 665
QY 763 QDPKRYGILYNDNEY-----LLYODWYLLS-----SKVNSPEDYFYVYKG 806
Db 666 QD-----LYDSKSDISIKLREKSSRIIAERFKLSNTLIDLKAKENDQKRFIDLQ 719
QY 807 -----RANDMDYGG-----SVLYTRSAVLPEPSIPELOTAQKVRDENTFIKT 833
Db 720 TILKODSKTHETINEYSCSKSLIYETELNLKE-----QKRLVLEKLNQEL 770
QY 853 NTCGPE-----PLVERLEKKEVEGERTITKEVELEEEVEKVRKXETLFSKL 901
Db 771 NKLSPEKDSIRIMYLOLQLOKEREDELLETBRSQCKRIDELEDLSEK--KET---SQK 826
QY 902 FEGKEKELORBE-----ENFLRELSKEEMVDLDGLKMEATVEKLFGRALPIRLMAVA 954
Db 827 DHITKOLEEDNNSNIEMQKIEAKKDYVSIVTSVDSQOTIEKLOKVKSKLEKEIE-- 885
QY 955 THCFSPCHRIREFS-----SDDGICRLQITKRNIGFLKILPIQISADLRTG 1006
Db 885 -----EDKIRLHTYANVDEITINDSL--RKELEKSKINLMDAVSQIR--EYKOLYET- 933
QY 1007 GSSRPPLSAFSGSGKGFIDVPLPSKNELEKELAPLLKLVGLACAFIYPSADAVDA 1066
Db 933 ---TSOSLOQOTNSKLDESKDFT---NOIKNLT-----DEKTS 964
QY 1067 LKTCACILKGCRIELAKCIANPACANVACLOTNNRPDETCQICGLFEN-----SV 1121
Db 965 LEDKISILKEQWENL-----NNELDLQKKMEKADFRKRISILONNKEVEAV 1014
QY 1122 VDEFNCAVSRKKCVPRKSDIGEPAPDPVSYLVONFNISDFNKNWITSLNTPDFADFC 1181
Db 1015 KSEY-----ESKLSKINDL---DOQITTYANTQNNYEDOLEKHAHVSKITISELRE 1062
QY 1182 OLHEFTEEDNKLGVNISWIKITIDSGFTTSVAVQKFOVDPNOGVLYVHNDNEYLAHQD 1241
Db 1063 QLHTY-----KQOVKTLN---LSRDOLENALEKNEKS---WSSQSKSLLEQD 1104
QY 1242 WYLLSKIKENKP-----DYFVYVYRGRNDAMQSYG-----AVYVYSSVLPNSIIP 1289
Db 1105 ---LSNSRIEDLSQNKLLYDOIOITYAADKREVNNTNPGGLNILLILRLRROLDTKYV 1162
QY 1290 ELKRAKKSIGRDFST-----IRTDNCTGPEPALVERIEKTEVEEGERTIYKEVEIE 1341
Db 1163 VAEEDAKMLKQISLMDVDELQDARTKDNS-----PVEK---ENHSSIIIOOHDIM 1210
QY 1342 EYVEKEVEKVRTEMILFORLAGFNEKQDEENFVPELSKEMEFDEIKMEASEVE 1399
Db 1211 EKL--NOJNLRESNITL-----RNEI--ENNKNKKELQSE---LKLQONVAPIE 1255

```

RESULT 9
DMD_CHICK

```

ID DMD_CHICK STANDARD; PRT: 3660 AA.
AC P11533;
DT 01-OCT-1989 (REL 12, CREATED)
DT 01-OCT-1989 (REL 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL 32, LAST ANNOTATION UPDATE)
DE DYSTROPHIN.
GN DMD.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX LEMAIRE C., HEILIG R., MANDEL J.L.;
RX MEDLINE: 89098331.
RT "Nucleotide sequence of chicken dystrophin cDNA.";
RL NUCLEIC ACIDS RES. 16:11815-11815(1988).
[2]
RP SEQUENCE FROM N.A.
RX TISSUE-MUSCLE;
RX MEDLINE: 89210800.
RA LEMAIRE C., HEILIG R., MANDEL J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal coding and 3' untranslated regions between man and chicken.";
RL EMO J. 7:4157-4162(1988).
CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE PLASMA MEMBRANE.
CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIBRIN, APP-120, APP-180, OR BETA-FODRIN).
CC -1- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL: X13369; G63370; -.
CC PIR: S02041; S02041.
CC PROSITE: PS00019; ACTININ_1; 1.
CC PROSITE: PS00020; ACTININ_2; 1.
CC PROSITE: PS01159; WW_DOMAIN_1; 1.
CC PROSITE: PS00020; WW_DOMAIN_2; 1.
CC PFAM: PF00307; actinin-binding; 1.
CC PFAM: PF00397; WW_tsp5_wmp; 1.
CC PFAM: PF00435; Spectrin; 22.
CC DR HSP: P00569; 42; 1.
CC DR HSP: 001082; 1432.
CC
CC STRUCTURAL PROTEIN; ACTIN-BINDING; CALCIUM-BINDING; CYTOSKELETON;
KW REPEAT.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 300 3000 26 SPECTRIN-LIKE REPEATS.
FT DOMAIN 3052 3085 WW DOMAIN.
FT DOMAIN 3086 3357 CYS-RICH.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1859 1869 O -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; AF61A205 CRC32;

```

Query Match 1.8%; Score 135; DB 1; Length 3660;

Best Local Similarity 17.6%; Pred. No. 12;

Matches 245; Conservative 196; Mismatches 482; Indels 472; Gaps 64;

```

QY 71 KDKSQICSIDTSEPELORPDKRGW--TLI--LEKQPOF-----IQALIVLC 115
Db 1923 KKKEDLNVAWNQJAEKSKDGAANAVEPTLVQSKRDFESKFAFRRLNAQIQVLE 1982
QY 116 TFIYIPRYDAVDALKTACCLKEKRIELAKCIANPACANVACLOTNNRPDETCQIK 175
Db 1983 TTFVWTESMIVETIYVPSYLAEL-IQLQALSEVEERLNSVLDQAKD-----C 2030

```

```

OY 1064 VALKTCACLLKGCRIELAKIANPACANVACLOTQNNPDETCQIR-----1
Dd 2927 DEALER-----1
OY 1113 GGDLEFNSVYDERNECAVSRKKCVRR-----SDIGEPPADPPSVLYONFNIS---DEN 1
Dd 2960 VGDLLIISLDQHLIEKRVYRAMVPLKEKRVQVNEIHAHFRAP-  
PDIIQSPSYLTSLGDELN 3
OY 1164 GRWYITSGNLPTDPA2CD2COHEFTE-EDNKLNGNISWRKTLIDSEFFLRSAVQ---K2V 1
Dd 3019 TRKRV---LQVALIDERIQLHAAHDEPPT-----SOHFLTTSVQPWERA 30
OY 1220 QDPNQPVLKHNHNE 1234
Dd 3062 ISPNKVPYYIYNHETQ 3076
RESULT 10
MISP_HUMAN
ID_MISP_HUMAN STANDARD: PRT; 1937 AA.
AC P13535: O14910.
DT 01-JAN-1990 (REL. 13, CREATED)
DT 13-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN, PERINATAL SKELETAL MUSCLE.
MY8.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 90323631.
RA KARSH-MIRZACHI I., FEGHALI R., SHORS T.B. JR., LEINWAND L.A.;
RT "Generation of a full-length human perinatal myosin  
heavy-chain-encoding cDNA."
RL GENE 89:289-294(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 95324556.
RA JULIAN E.H., KELLY A.M., POMPIDOU A.J., HOFFMAN R., SCHIAFFINO S.,
RA STEWMAN H.H., RUBINSTEIN N.A.;
RT "Characterization of a human perinatal myosin heavy-chain  
transcript."
RL EUR. J. BIOCHEM. 230:1001-1006(1995).
RN [3]
RP SEQUENCE OF 502-1937 FROM N.A.
RX TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 90235862.
RA BOBER E., BUCHERGER-SEIDL A., BRAUN T., SINGH S., GOEDDE H.W.,
RA ARNOLD H.H.;
RT "Identification of three developmentally controlled isoforms of human  
myosin heavy chains."
RL EUR. J. BIOCHEM. 189:55-65(1990).
RN [4]
RP SEQUENCE OF 860-1937 FROM N.A.
RX MEDLINE; 89234168.
RA FEGHALI R., LEINWAND L.A.;
RT "Molecular genetic characterization of a developmentally regulated  
human perinatal myosin heavy chain."
RL J. CELL BIOL. 108:1791-1797(1989).
CC -1 FUNCTION: MUSCLE CONTRACTION.
CC -1 SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1 SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1 DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES  
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1 PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
AKTYLATED AND ARE REQUIRED FOR

```


QY 1165 KNYITSGNPTDAFCOLHEHTEGDKLVGNISMRKLTGDSGFTRSAVQKVFDPNQ 1224
 Db 1413 K---CASLEKTKORLQNEVEDMLDVE-----RSNACALDKKO 1449
 QY 1225 PGLVNLHNEVLIHODDMYILSSKIENKRPDIYFVYRGNDAMDYGAVYITSSVLP 1284
 Db 1450 ---RNFQVLSSEKQKKEETQALFLASQKE-----SRSLSTELFKVKNVEESL--- 1496
 QY 1285 NSIIPLEKAKSIGRDFSTFIRNDNCGEPALVRIKTYVEEGERIYVEVEPIEEV 1344
 Db 1496 -DOLETLRREKKNQOEISDLTEOIALAGGQKHLEKIKVOEKECELOALEEAEASL 1554
 QY 1345 EKEVEKVGRTMELFO-----RLAEGFNEKODEENFVELS----- 1382
 Db 1555 EHEGKTLRIQLELNOVKSEVDKRIAEKDEIDQLKRHTVAVELMOSTLDAETRSMDA 1614
 QY 1382 ---KEENE-FLDEIKKAESEVERLFGKAL 1406
 Db 1615 LRVKMKMEGDLNEMEIQLNHNRLAESI 1643

RESULT 11

MSPL_PLAFK STANDARD: PRT: 1630 AA.

AC 13-AUG-1987 (REL. 05, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-OCT-1995 (REL. 34, LAST ANNOTATION UPDATE)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
 DE (PMMSA) (P190).
 GN MSP-1.
 OS PLASMODIUM FALCIPARUM (ISOLATE K1 / THAILAND).
 OC EUKARYOTA: ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86136024.
 RA MACKAI M., GOMAN M., BONE N., HYDE J.E., SCAITE J., CERTA U.,
 RA STUNNENBERG H., BUARD H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 RT Plasmodium falciparum merzoites: studies at the genetic level.",
 RN EMBO J. 4:3823-3829(1985).
 [2]
 RP REVISIONS. SEQUENCE FROM N.A.
 RA PAN W., TOLLE R., BUARD H.;
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KD, 42
 CC KD AND 19 KD ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC EMBL: X03371: G929798: .
 DR PIR: A25120: SAZOK1.
 DR PIR: PR00008: EGF: 1.
 KW MALARIA; MERZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
 KM TRANSMEMBRANE; GPI-ANCHOR.
 FT SIGNAL 1 19
 FT CHAIN 20 1630 POTENTIAL.
 FT DOMAIN 67 84 MERZOITE SURFACE PROTEIN 1.
 FT TRANSMEM 1614 1630 TRIPEPTIDE SG(TP) REPEAT.
 FT CARBOHYD 97 97 MEMBRANE ANCHOR.
 FT CARBOHYD 259 289 POTENTIAL.
 FT CARBOHYD 755 755 POTENTIAL.
 FT CARBOHYD 759 759 POTENTIAL.
 FT CARBOHYD 774 774 POTENTIAL.

FT CARBOHYD 835 835 POTENTIAL.
 FT CARBOHYD 911 911 POTENTIAL.
 FT CARBOHYD 955 955 POTENTIAL.
 FT CARBOHYD 1049 1049 POTENTIAL.
 FT CARBOHYD 1156 1156 POTENTIAL.
 FT CARBOHYD 1165 1165 POTENTIAL.
 FT CARBOHYD 1436 1436 POTENTIAL.
 FT CARBOHYD 1517 1517 POTENTIAL.
 SQ SEQUENCE 1630 AA, 187289 MW, DD2F8628 CRC32;

Query Match 1.8%; Score 134; DB 1; Length 1630;
 Best Local Similarity 18.0%; Pred. No. 4.7;
 Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

QY 184 VDOFNECAVSRKKVCRKSDVEFFVPPDRAVONFNMKDPSGKWTISGLNPTFAFC 243
 Db 240 IENINELIESKTKIDK-----NKNATKEEKKKLYQAQYDLS-----IYKK 281
 QY 244 OLHEFHMENDKLGVNLTWRKIKTLDSGFTRSAVQTF--VODP-----DLGALYNHNE 295
 Db 282 QLEEAH---NLISYLERIDITLKNENIKELDKINEINPPANGNTPTMLDKNKK 337
 QY 296 FLHYDDWYILSSQENKPDY-----IFVYRGNDAMDYGGSVITRSPITLESII 349
 Db 338 IEHEKEIKELAKTIKRNIDSLFTDPLELEYLREKKNID--ISAKYETKESTEPNE-Y 394
 QY 350 PVLQKAASVGRDNNFTTNSCGEPPLVERLEKTAEGEKLL-----KEAVEIEE 404
 Db 395 PNGVYPLSTY-NDINNALNELNSFG--DLINPDTKPSKNITTDERRKFTNEIKER 450
 QY 405 VEKEVEKVRDTEMTLQRLLEGFKELQODEENFVELSKEKEILNEL-----OMEAT 457
 Db 451 IKIKKKRI-----ESPKASYEDRSKSLMDITKEKELNETYDSKFPNNIDLT 498
 QY 458 EVELKFGK-ALPIKRLRALAPHSNFLANHETIYYGSKLPGRKFPNGWEDY-FGSI 514
 Db 499 NFEKMKRKRISYKVEL-----THNNTFASYENSKHNL-EKLTALKY---MEDYSLRNI 549
 QY 515 VYAKICSSRRIPRPFKPSRICGDSRGLOTFSGHNSPASHISNOVNPKNQNSGC-- 572
 Db 550 VVEREL-----KYYK-----NLISKINET--ETLVENIKKDEQLFEK 586
 QY 572 ---KPKDVALMWMKQGFAPKTAIVAFILSVASKADAVADAKTCTCLKEERLE- 625
 Db 587 KITKDNKPDEILTV---SDIVQVQKVLIM-----NKIDELAKTQILIKVNEVKH 636
 QY 625 ---LAKCISN---PACAA-----NVACLOQNNRPDET 651
 Db 637 NIHVNSYKQENKQEPYLLIVLKKEIDKLKVPKVESLINEKKNITREGOSDSEPT 696
 QY 652 EGOI-----KCGDLFENSVDVEFNCAVSRKKCVPRKSDGDPVDPSTVLY-- 700
 Db 697 EGEITGQATTRPGQAGSALFG--DSVQAQOEOKQAP-----PVPVVPKAKQVP 747
 QY 700 ---KIDMDFSGKWFITRGLNPTFARDCOLHEHTEGDKLVGNISMRKLTGDSGFTR 745
 Db 748 TPAPVANKTKENVS KDYLE-----KLYEFL-----NTSYIC 779
 QY 746 RTPDGGFFTRSAVQKVFQDPKYPGILYNHD--NELLYQDDWYILSSKVEN--SPED 798
 Db 780 H-----KTLVSHSTNNEKILKO--YKITEEESKSSODPLD 815
 QY 799 YIF-----YVYKRGNDAMDYGGSV-----LYTRSAVL-----PESIIPELOTAQAQ 839
 Db 816 LLEFNINNIPIVAY---SMFSLNNSLSQLFMEIYEKEMCNLYKLKDKDKIKNLEAK 871
 QY 840 KYGRDNTFIKTDNCGEP--PLVERLEKKEVEGERI-----LIKVEEII----- 884
 Db 872 KV---STSVKTLSSSSQPLSTLPQDKPEVSANDTSHSTINNSIKLFLFNISLGKNN 927
 QY 884 ---EEVEKVRDKEVTFLESKLFGFKELQODEENFVERLEKTAEGEKLL----- 931

[illegible]

[illegible]

```

Db 567 NKGKQEAH-----ESLVHAGTVINILGW-----LQKNKBDPLNEVGLVLYOKSSILK 615
OY 359 VGRDNNFTITDNGSGPPPLVERLEKTAEGEKKLLIEAVEI---EEFEVEEVKVDTE 416
Db 616 LSNLFANT-----AGADAP--VDKGGKAKKGSF---QIVSVLHREKLNKLMNLTST- 665
OY 417 MTLFOLLEGEKELQODENFVRELSEKEKE-----ILNELMEAT--EVEKELGRA 466
Db 665 -----HPHFVRCIIPNETKSPGMNPLVMHQLRCNGVLEGIRICRG 707
OY 467 LP-----IRKLMALAPHSNPLANEITIKYVGSKLPQHKRFSWGMEDYFESSIV 515
Db 708 FPNRLLYDFRQRYRILNPAIPEGQFIDSRKGAEKLLSSDIDHNOYKFGHTVFFERAG 767
OY 516 VAKICSSRIIPRYFKSPRICGGLDSRGLQFSKGNHLSPAHS---INQVPRKNSGCK 572
Db 768 LIGLLEHRDERLSITIRI--QAOSRGLSRMEFKLLERDSLVIYQWNI--RAFMGVK 824
OY 573 -FP-----KVALMVMKMGQFAKTAIVAFILSVASKADAVDAKLT 613
Db 825 NMPWMLKLYFKIKPLKSAETKEKMATK--EEFGR-----VKDALEK 864
OY 614 CTCLLEKRELEACISNPACANVACLOTGNNRPDETECOIKGDFENSV-----VDE 668
Db 865 SEARRKELEKVVSLQF-----KNDLOQVQAEODNLADAEERDOLIKNKIQLEAKYKE 920
OY 669 FNE-----CAVSRK--KCVPKSDVGDGFPVPSVLYOKFDMKDFSGKMFIT 713
Db 921 MTERLEDEENNAELTAKKRLIEDCSSELKRDIDDLITLAKVEKD----- 968
OY 714 RGLNTPDADFQOLHEFTEENKLVNLSWRIRTPDGGFTSRSAVOKEFVDPKPYGILYN 773
Db 968 -----HATENK--YKNLTSEMAGLD-----ETIAKLTKEKKAQLOEAHQ 1003
OY 774 HDNELLYQDDMY--ILSSKYE-----NSPEDIYVYIKGRND---AMDGGSVLYITRS 823
Db 1004 QALDLOAEERKVTNLTLSKVKYLEQOVDDLEGSLEQEKRVMDERAKRKLEGLDKLTQE 1063
OY 824 AVLPEIIPELQTAOKYG-----RDF-----NPFIKIDNCGPE-----PRIVERL 865
Db 1064 STM-----DLENDKQQLDEKIKKKDFELNALNARIEDQALGSLQKKELOARIEEL 1117
OY 866 EKKVEEGERT-----IIEKEEIEEVEK----- 890
Db 1118 EEL--EABRTARAKYKELRSDLSRELEISERLEAGATSVQIEMNKKRAEPQKMRD 1116
OY 890 -----VRDKEVILFSKLFEGEKELODENFLEISKEEMDYLDGKMEAT 935
Db 1177 LEELATLOHEATMAALRKHADSVALGEOINDLQVOKOLEKSEKFEKLEIDD-----VTS 1232
OY 936 EYEKLEFGALPIRKLMATAVHCFSPCHDRIFESSDGIIGLITRKIRINGTELKILP 995
Db 1233 NMEQITKAKANLEK-----CRT--LEDQMEHRS-----KAETQRSVNDL----- 1273
OY 996 PIQASDLRTGGSSRPL-----SAFRSGFSKGIPTVPLPSKNEKLELAPLLKLGVLA 1052
Db 1273 TSQRAKLOENELSRQDEKALISQLTRG-----KLTYYQQLDDKRLQ----- 1319
OY 1053 CAFLLVPSADAVDAKTCACILKGRITELACIANPACANVACLOTGNNRPDETE--CQ 1110
Db 1319 -----EEEVKANNITLAHALQSAHND-----CDLREQYEETEAKAE 1355
OY 1111 IKC-----GDUFENSVDEFNECAVSRKCVPRKSDGGEFAPDPSPVLYQNFN 1158
Db 1356 LQCVUSKANSEVAQWTKRETDAIQRTLEBAKRLAQRQDAE----- 1402
OY 1159 ISDFNKWYITSGLNTPDADFQOLHEFTEENKLVNLSWRIRKTLDSGFTSRSAVOKE 1218
Db 1402 -----AVEAVNAKSSLEKTKHRLQNEJEDLMVD-----VERSNMAAA 1439
OY 1219 VQDPNPQGVLYHHDEYLYHQDDWYILSSKLENKPEDIYFYVYRGRNDAMDQYGGAVYT 1278

```

```

Db 1440 ALDKKO-----RNFDKILAEWKQKEESQSELESCKE-----ARSLST 1478
OY 1279 RSVSLPNSIIPLE-----KAASIGRDFSTFIPTDNTCGEPALVERIEKTVEEGERI 1332
Db 1479 ELFKLNAYEESSEHLETKRENKNQDEISDITQOLQSTGSHLEKIRQLOAEKME 1538
OY 1333 IYKEVEIEEVEKEVEKVGRTMPLFQ-----RLAEGNELKODENFVRELS----- 1382
Db 1539 LQALBEAASLEHEGNTILRAQLEFNQIKAEIERKLAEKDEDEQAKRNHLRVDSLOT 1598
OY 1382 -----KEEME-FLDEIKMEASEVEKLFQGA 1405
Db 1599 SUDAEIERSNEALRYVKRKMEGOLNEMEITQLSHANRMAEA 1638

RESULT 14
LEF_BACAN
ID LEF_BACAN STANDARD: PRT: 809 AA.
AC P15917;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).
GN LEF.
OS BACILLUS ANTHRACIS.
OC PLASMID PX01.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.
RX MEDLINE: 90034185.
RA BRAG T.S., ROBERTSON D.L.;
RT "Nucleotide sequence and analysis of the lethal factor gene (lef)
RL GENE 81:45-54(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA LOWE J.;
RL SUBMITTED (APR-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP ZINC-BINDING.
RX MEDLINE: 95154669.
RA KOCHT S.K., SCHIAVO G., MOCK M., MONTECUCO C.;
RT "Zinc content of the Bacillus anthracis lethal factor.";
RL FEMS MICROBIOL. LETT. 124:343-348(1994).
CC - FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. LEF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED
CC WITH PA, CAUSES DEATH. LEF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO
CC BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY
CC FACILITATING THE INTERNALIZATION OF LEF OR EF.
CC - SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT
CC PROTEINS, A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN
CC EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC - SUBCELLULAR LOCATION: SECRETED.
CC - SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B. ANTHRACIS EF
CC AND LEF.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC
CC METALLOPROTEASE).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M29081; G143144; -
DR EMBL: M30210; G143142; -
DR PIR: JQ0032; JQ0032.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW HYDROLASE; METALLOPROTEASE; ZINC; TOXIN; SIGNAL; REPEAT; PLASMID.

```

FT SIGNAL 1 33
 FT CHAIN 34 809 LETHAL FACTOR.
 FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
 FT DOMAIN 300 420 REPEATS.
 FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
 FT ACT_SITE 720 720 POTENTIAL.
 FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
 FT SEQUENCE 809 AA: 93786 MW: D81B6RBB CRC32:

Query Match 1.8%; Score 132; DB 1: Length 809;
 Best Local Similarity 17.2%; Pred. No. 2.4;
 Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 EENKLVGNISMRIRTPDGGFTTRSAVOKVOPKPGILYHND---NEYLYODDWY--I 787
 DB 105 EMYKRAIG--KIVYDGGITRHSISLEALSEDKRKIKDYLKDALHHRVYAKGEYEPV 161
 QY 788 LSSKVENSPEDYIFYYYKGRNDAMDGGYGVLYTRSAVLPESIIPELOTAOKVGRDENT 847
 DB 162 L---VIGSSEDIYVENTERKALN-----VYEEIGKILSRILSKINOPYOKFLDVLNT 209
 QY 848 FRTDNTG-----PEPLVERLEKKEVEGERTIIR----- 879
 DB 210 IKNASDSDGODLFTNOLKEHPTDFSVEFLQNSNEVOEFAKAFAYIEPOHRDVLQLY 269
 QY 879 -----EVELEEEVEKVRDKEVTLFSLKEGFRK----- 908
 DB 270 ABAENYNDKFEQENINLSLELKDORMLSRYEKWKIKQYOHWSDSLSEGRGLKRL 329
 QY 908 ---LORDEENFLRELSEKEMDVLIDGKMEATEV---EKLEGRALPI-----RKL 950
 DB 330 QPIEBRKDIIHSLSQEKELKRIQIDSSDPLSTEKEFLKQIDIRDSLSSEKEL 389
 QY 951 MAVVAHCFSPCHDRIRFRSSDDGIGRLGILTKRINGTFLKILPPTOSAD---LRFTG 1006
 DB 390 L-----NRIQVDSNP-----LSEK--EKEFLKLTADIDPYINORLODPTG 429
 QY 1007 GRSRPL-----SAFRSGFGKGFIDIVPLPSKNEMLKELIAPLILKLVG 1049
 DB 430 GLIDPSTILDYRKQYKRIQINIDALHQSISTIKYIKYLEMNNINNLITL----- 484
 QY 1050 VLAACFLVPSADAVADALKTACILKGRICELAKCIANPACANAYACLOTGNNRPDETEC 1109
 DB 484 -----GADLYDST----- 492
 QY 1110 QIKGDLFENSVD--FNECAVSRKKCVPRK---SDLGEPAPDPVLYVONFNISDENG 1164
 DB 492 -----DNIRKINGINEFEKKNRYSTISSNYMTAVDINERPALDNERLKWRIQISDPT 543
 QY 1165 KWTYISGLNPTDAFDQCLHEFHTEGDKLV--GNISWRIKTLDSGFTTRSAVOKFVODP 1222
 DB 544 AGYLENG-----KLILQRNIGLEINDVQ----- 567
 QY 1223 NOPGVLYHNDNYLYHODDWYLLSKTIENKPEPDYIFYVYRGANDAMDGYGAVVYTR--S 1280
 DB 567 ---IKOSEKYEYIRI--DAKVPRKSIDIKIOE---AQLNINOMNKAALDLPYTKLIT 617
 QY 1281 SVLPNSIIPLELEKAKSIGRDFSTIRTDNCGPEPALVERIEKVEEGE--RIIVKEV-- 1338
 DB 618 FVYHRRYASNIYESAYLLINENKNNIQSD-----LIKVTYINLVGNGREVFDTITL 669
 QY 1338 -----EIEEVEVEKEVKGRTMTLFORLAEGFNEIKODENFVRE----- 1380
 DB 670 PNIAEQYTHODEIYEQVHSHKGLYVESRSILHGPSKV--ELRNDSEGFIEHFGHADV 728
 QY 1380 ---LSKEEM-----FLDEIKHASEV 1398
 DB 729 AGYLLDKNOSDLVYNSKKFLIDLEKEGSLN 758

ID Y109_YEAST STANDARD: PRT: 1679 AA.
 AC P40457;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE HYPOTHEICAL 195.1 KD PROTEIN IN DNA3-DBII INTERGENIC REGION.
 GN Y1149C.
 OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACHAROMYCETALES;
 OC SACHAROMYCETACEAE; SACHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN-S288C / AB972;
 RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
 RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
 RA GUNTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
 RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PARSON D.,
 RA RAJANDRAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
 RA WALSH S.V., WHITEHEAD S.;
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC DR EMBL; 247047; G763197;
 CC EMBL; 238059; G557774;
 CC PIR: S48385; S48385
 CC KW HYPOTHEICAL PROTEIN.
 SQ SEQUENCE 1679 AA: 195141 MW: 5897CD94 CRC32:

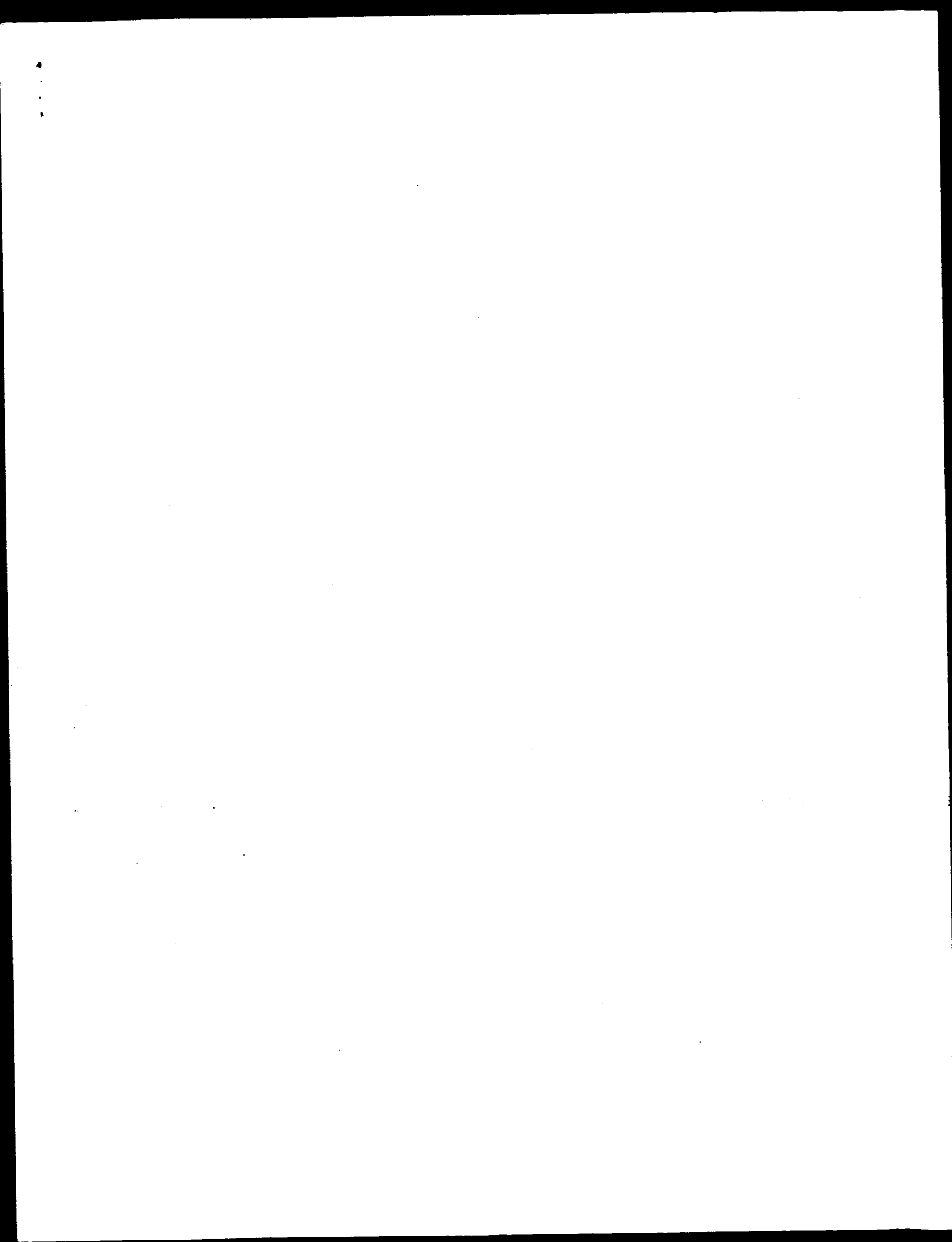
Query Match 1.7%; Score 130; DB 1: Length 1679;
 Best Local Similarity 17.9%; Pred. No. 8.3;
 Matches 210; Conservative 188; Mismatches 396; Indels 376; Gaps 59;

QY 379 LVRLD-----KTAESGKLLIKAEVIEEVEKVEKVDTEMLTFORLLEKFEKLQ 431
 DB 536 IADKLENYGKODKTLQKVENOTIKKAKDAIYLE--NINAKMETRINILRRRDSYKLLA 594
 QY 432 QDEENFVR-----ELSEEKEIILNLOMTEVEKLEFGRALPIRKIR-- 474
 DB 595 STEENANNTSVYSMAAREKIRLELAEISSTYVE---NSAIIQNLKELLIIYKKSQC 650
 QY 474 ---MALAPHSNF--LAN-----HETIKYVGSKLPGHKRFSGWEDYFGSIYAKICSS 522
 DB 651 KKKTLEDENFKGLAKKEXRMLLEAIDHL---KALEKKQSM----- 691
 QY 523 RIRPRIYR-KSPRICGGLDSRLQLEFSGKHNLSPASHINON--VPKNS-----GCKF 573
 DB 691 ---VPSTIHYEKERASTELQSRIKISL--EYELISLKKRETSFIPTKESLIRDEFOCKE 747
 QY 574 PKDVALMVE-----KMGOPAKTAIYAI-----FLVSASADAVADAKT 613
 DB 748 KKLQRLKLESELSHKNKMDSSKEGY--KAKIKELENNLERLSJDQSKIQTETESIRS 806
 QY 614 CTCLLEKRLLELAKCISNPACANAYACLOTGNNRPDETECOIKGCDLFENSVDENFCA 673
 DB 807 C---KDSQLKMAQ-----NTIDPTEMKKK-----SLTFELSKKE 837
 QY 674 VSKKKVPRKSDVGFPPVDPVLYOKEDMKDFSGKWTITG-----LNPV---FDAFD 724
 DB 838 TITELK---SSELENL---DKELRKTFOYK-----FLDQNSDASTLEPLTKLELEIQ 885
 QY 725 COLHEFTEENKLVGNLSWRIRTPDGGFTTRSAVOKFVODKPPGIIYNHNEYLYLQDD 784
 DB 886 VOLKQANS-----QIQAYEETISSNENALI----- 911
 QY 785 WYLLSKVENSPEDYIFYYYKGRNDAM-----DGYGGSVLYTRSAVLPESIIPELO 835

RESULT 15
 Y109_YEAST

```
Db 911 --ELKNEIAKTEKNEYDAKIELEKKEKWARREDLSRLGELGEI-----RALQPKLK 959
QY 836 TAAQKVRGDEFTFKTNTGCPPEPLVERLEKKEVEGER--TTI-----KEVEEIEEVE 888
Db 960 EGALH-----FVQOSEKIRNE---VERIQKIEKIEKSTIVQJCKRREMSQYOSTMK 1009
QY 889 KVRD-----KEVT-LFSKLEGEKELQDEENFLRELKSEEMVDVLGK 931
Db 1010 ENKDLSELVRLKEDADCCQAEILTKSSLYSAQDLLDKHERKWMERKADYERELISNIE 1069
QY 932 M-EATEVEKLEFGALPIRKLMAVATH-----CETSPCHDIRFTSSDDGIGRLG 979
Db 1070 QTESLRE---NSVLEKVEDDTAANNQDKDHUKLVLSFSNLHER----- 1112
QY 980 ITRKRINGITFLKILPPIQADLR--TTGGRSSRPISAFRSGFSGIFDIIVPLPSKNEIKE 1038
Db 1112 -----NSLETKLFTCKRELAFVQKNDLSLEKTIINDLQRTQITLSE-KE 1152
QY 1039 LTFAPLLKLVLGVLACAFILVPSADAVDALKTACCLIKGRIELAKCIANPACANVACIQ 1098
Db 1153 -----YQSAVLIIDEFKDITKEVTQVNLKENNALIQKSLKN-VTEKNREIYK 1199
QY 1099 TCNNRPDETECOIKCGDLFENSVDNEFCAY-----SRKKCVPRKSDLG--EFPAPDPS 1151
Db 1200 QLNDRQOEISRLQR--DLIQTKQOYSINSNKILVYSEMEOCKORYQDLQSOQKDAQKRD 1257
QY 1152 VLVQNFNIDFNGKWTITSGLNPTPDAPDCQLHEFHTEGDNKLVGNISWRIKTLDSGFFT 1211
Db 1258 IEKLINEISDLKG--LSSAENANADL-----ENKFN-----RLK----- 1291
QY 1212 RSAVQKRVQDPNQPGVLYVHNDEY-----LHYDDWYI--LSEKI--ENKPPEDI 1257
Db 1291 KQAHKELDASKKQOQALINELNELKAIKDLBODLHFENAKVIDLDTKLKAHELQSEDV- 1350
QY 1258 FVYIYGRNDAMDYGGA VYTRSSVLPNSIIPLEKAKASIGRDSFTFIRTDNCGPEPA 1317
Db 1350 -----SRDHEKDTY-----RTIMEEIE-----SLKRELQIF-----KTANSSD 1383
QY 1318 IVERIEKTYE--EGERIIVKEVEIEEVEKEVEKVGRTENTLFORLAEGFNE--LKODEEN 1375
Db 1384 AFEKLVNNEKEKDRIDERTKEFEKKLQETLNKSTSEAEYSKDIETLKKEWLKEYEDE 1443
QY 1376 FVRELKSEMEFLDEIKMEASE--VEKLFQK 1404
Db 1444 TLRIKAEAEENLKKRIRLPSEERIQIKIISK 1473
```

Search completed: October 14, 1999, 03:57:21
Job time: 2354 sec



GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 1999, 22:33:02 : Search time 44.23 Seconds
(without alignments)
1964.721 Million cell updates/sec

Title: US-09-075-375-2

Perfect score: 7495

Sequence: 1 MALSLHTVFLCKEALNLXA.....MEASEVEKLFGRKALPIKRV 1412

Scoring table: BIOSUM62

Searched: 201082 seqs, 61543640 residues

Database :

SPTREMBL_10.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mmc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2357	34.1	478	10	Q40593	Q40593 nicotiana t
2	2508	33.5	473	10	Q40251	Q40251 lactuca sat
3	2430	32.4	462	10	Q39249	Q39249 arabidopsis
4	194.5	2.6	2269	5	Q26223	Q26223 plasmodium
5	167.5	2.2	2401	5	Q26216	Q26216 plasmodium
6	136	2.1	1365	2	Q49525	Q49525 mycoplasma
7	156	2.1	1933	13	Q90337	Q90337 cyprinus ca
8	155.5	2.1	886	1	Q29230	Q29230 archaeoglob
9	151	2.0	1939	5	Q25662	Q25662 plasmodium
10	147	2.0	3113	4	Q13246	Q13246 homo sapien
11	146	1.9	2748	3	Q03767	Q03767 saccharomyc
12	144.5	1.9	839	5	Q26024	Q26024 plasmodium
13	144	1.9	2166	2	Q51465	Q51465 borrelia bu
14	143.5	1.9	1558	5	Q96275	Q96275 plasmodium
15	142.5	1.9	1819	2	Q92LV0	Q92LV0 helicobacte
16	142.5	1.9	1199	5	P91349	P91349 caenorhabdi
17	142.5	1.9	5105	5	Q61201	Q61201 caenorhabdi
18	141	1.9	1302	2	Q65583	Q65583 aquifex aeo
19	140.5	1.9	1302	2	Q49547	Q49547 mycoplasma
20	140	1.9	800	1	Q59066	Q59066 methanococ
21	139.5	1.9	1025	3	Q12176	Q12176 saccharomyc
22	137	1.8	1109	6	Q00756	Q00756 oryctolagus
23	137	1.8	1676	10	Q23332	Q23332 arabidopsis
24	136.5	1.8	3724	5	Q77320	Q77320 plasmodium
25	136	1.8	1127	12	Q9YVT6	Q9YVT6 melanoplus
26	135.5	1.8	1312	4	Q92878	Q92878 homo sapien
27	134.5	1.8	946	2	P70888	P70888 bacteroides
28	134	1.8	978	2	O67124	O67124 aquifex aeo
29	134	1.8	620	4	O43663	O43663 homo sapien

30	134	1.8	2261	5	O62175	O62175 caenorhabdi
31	134	1.8	1935	13	Q90339	Q90339 cyprinus ca
32	133.5	1.8	1318	4	O43254	O43254 homo sapien
33	133	1.8	1624	5	P91121	P91121 caenorhabdi
34	133	1.8	1979	5	O96133	O96133 plasmodium
35	132.5	1.8	1156	2	O68878	O68878 aquifex aeo
36	131	1.7	1156	1	O28714	O28714 archaeoglob
37	130.5	1.7	962	2	Q49546	Q49546 mycoplasma
38	130	1.7	1938	6	Q28641	Q28641 oryctolagus
39	130	1.7	1931	13	Q42352	Q42352 cyprinus ca
40	129.5	1.7	993	2	Q46673	Q46673 lactococcus
41	129.5	1.7	1992	13	Q04834	Q04834 xenopus lae
42	129.5	1.7	2954	13	Q42263	Q42263 xenopus lae
43	129	1.7	1179	1	O59462	O59462 pyrococcus
44	129	1.7	1201	5	O18392	O18392 dirosophila
45	129	1.7	2919	12	Q85431	Q85431 rice stripe

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	AA
Q40593	Q40593			
AC	Q40593	01-NOV-1996 (TRENBL)	01	Created
DT	01-NOV-1996 (TRENBL)	01	Last sequence update	
DT	01-JAN-1999 (TRENBL)	09	Last annotation update	
DE	VIOXANTHIN DE-EPOXIDASE PRECURSOR			
GN	TYDEL			
OS	Nicotiana tabacum (Common tobacco)			
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;			
OC	euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:			
OC	core eudicots: Asteridae: euasterids I: Solanales: Solanaceae;			
OC	Nicotiana			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-XANTH: TISSUE-LEAF;			
RA	BIGOS R.C., YAMAMOTO H.Y.;			
RL	Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: U34817; AF050031.1; .			
DR	MEDLINE: 9222; NIDA:Vdel.1.			
KW	Transit peptide.			
FT	TRANSIT 1 134			POTENTIAL.
FT	CHAIN 135 478			POTENTIAL.
SEQUENCE	478 A; 54561 MW; 9882AD42 CRC32;			

Query Match 34.1% Score 2557; DB 10; Length 478;

Best Local Similarity 100.0% Pred. No. 2.5e-139;

Matches 477; conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	474	MALAPHSNLANHETIKYVGSKLPGHKRFSWGEDYFGSIIVAKTSSRIPIREKSP	533
DB	1	MALAPHSNLANHETIKYVGSKLPGHKRFSWGEDYFGSIIVAKTSSRIPIREKSP	60
QY	534	RICCGLDNRGLQFSGHKNHNSPAHSTINONVPGKNSGCKTPKVOALMWKMOFAKTAI	593
DB	61	RICCGLDNRGLQFSGHKNHNSPAHSTINONVPGKNSGCKTPKVOALMWKMOFAKTAI	120
QY	594	VAIFLTVASRVDVADALKTCTCLKECRLELACISNPACAAVACLOTNNRPDETEC	653
DB	121	VAIFLTVASRVDVADALKTCTCLKECRLELACISNPACAAVACLOTNNRPDETEC	180
QY	654	QIKCGDLFENSVDENFENCAVSRKCVPRKSDVDFPVPDPSVLVOKFDMKDSGKWFIT	713
DB	181	QIKCGDLFENSVDENFENCAVSRKCVPRKSDVDFPVPDPSVLVOKFDMKDSGKWFIT	240
QY	714	RGUNTFPAFRCQLHEFTTEENKLVGNLSRIIRPDGFFTRSAVQKVFODPKYPGILYN	773
DB	241	RGUNTFPAFRCQLHEFTTEENKLVGNLSRIIRPDGFFTRSAVQKVFODPKYPGILYN	300
QY	774	HDNEVLTVQDDNVLTKSVKSPEDYIFVYKGRNDAMDYGGSVLTVRAVLPESTIPE	833

DB 301 HONREYLLYDDDWYLLSSKVENSPEDYIFYYKGRNDAMGYGSVLYTRSAVLPSSTIPE 360
QY 834 LOTAOKVGRDNTFTKIDNTGCGPEPLVERLEKKEVEGERTIIEVEEIEEVEKVRDK 893
DB 361 LOTAOKVGRDNTFTKIDNTGCGPEPLVERLEKKEVEGERTIIEVEEIEEVEKVRDK 420
QY 894 EYTLFSEKLEGEFKELOREDENFLRELSKEMDVLDGLKMEATEVKLRALPIKRL 950
DB 421 EYTLFSEKLEGEFKELOREDENFLRELSKEMDVLDGLKMEATEVKLRALPIKRL 477
RESULT 2
Q40251 PRELIMINARY: PRT: 473 AA.
ID 040251
AC 040251
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMblrel. 09, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN VDEL.
OS Lactuca sativa (Garden lettuce).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids II; Asterales; Asteraceae;
OC Lactuca.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ROMAINE;
RX MEDLINE: 96270536.
RA BUGOS R.C., YAMAMOTO H.Y.;
RT "Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
and expression in *Escherichia coli*.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:6320-6325(1996).
DR EMBL: U31462; AAC49373.1; -
DR MENDEL: 8691; LACsa.Vdel;1.
KW Transist peptide.
FT CHAIN 1 125 POTENTIAL.
FT CHAIN 126 473 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 473 AA: 54447 MW: 102E7001 CRC32:
Query Match 33.5%; Score 2508; DB 10; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.6e-136;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MALSHYFELCKEALNTLYARSPCNERHRSOGPTNTIMKTSNNGYNSFRLFTSYK 60
DB 1 MALSHYFELCKEALNTLYARSPCNERHRSOGPTNTIMKTSNNGYNSFRLFTSYK 60
QY 61 TSFSDSSHCKDKSQICSDTSFEIORFDLKRGMTLILEKOWROFOLAIVLCFTVIY 120
DB 61 TSFSDSSHCKDKSQICSDTSFEIORFDLKRGMTLILEKOWROFOLAIVLCFTVIY 120
QY 121 PRVAVALKTCACCLIEKRETEIAKCIANPSCANVACLOTNNRPDETCQIKGDLFE 180
DB 121 PRVAVALKTCACCLIEKRETEIAKCIANPSCANVACLOTNNRPDETCQIKGDLFE 180
QY 181 NSVVDQNECAVSRKRCVPRKSDVGEFPVDRNAVONFNMKDPSGKWTITSGINPTFDA 240
DB 181 NSVVDQNECAVSRKRCVPRKSDVGEFPVDRNAVONFNMKDPSGKWTITSGINPTFDA 240
QY 241 FDCOLHEFHENDKLVNLTWRIKTLTLDGFFTRSAVQTFVQDDPLPGALYNHNEFLHYQ 300
DB 241 FDCOLHEFHENDKLVNLTWRIKTLTLDGFFTRSAVQTFVQDDPLPGALYNHNEFLHYQ 300
QY 301 DDWYLLSSQJENKPDYIFYYKGRNDAMGYGSVLYTRSAVLPSSTIPE 360
DB 301 DDWYLLSSQJENKPDYIFYYKGRNDAMGYGSVLYTRSAVLPSSTIPE 360
QY 361 RDNENFTTJNSCGPEPLVERLEKKEVEGERTIIEVEEIEEVEKVRDK 420
DB 361 RDNENFTTJNSCGPEPLVERLEKKEVEGERTIIEVEEIEEVEKVRDK 420

QY 421 ORLLEGEFKELODENFVRELSEKEELINELQMEATEVKLRALPIKRL 473
DB 421 ORLLEGEFKELODENFVRELSEKEELINELQMEATEVKLRALPIKRL 473
RESULT 3
Q39249 PRELIMINARY: PRT: 462 AA.
ID 039249
AC 039249
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMblrel. 09, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN AVDEL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA BUGOS R.C., YAMAMOTO H.Y.;
RT Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SHINN P., BUEHLER E., DEMAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OJI O., SHEN Y.K., TORIMI M.,
RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: U44133; AAC50032.1; -
DR EMBL: AC003981; AAC14029.1; -
DR MENDEL: 6341; ARath.Vdel;1.
KW Signal.
FT CHAIN 1 113 POTENTIAL.
FT CHAIN 114 462 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 462 AA: 52017 MW: 37FB33BA CRC32:
Query Match 32.4%; Score 2430; DB 10; Length 462;
Best Local Similarity 100.0%; Pred. No. 4.5e-132;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 951 MAVATHCTSPCHDIRIFRSSDDGIGRLGTRKRINGFLKILPPIQSADLRTTGRSS 1010
DB 1 MAVATHCTSPCHDIRIFRSSDDGIGRLGTRKRINGFLKILPPIQSADLRTTGRSS 1010
QY 1011 RPLSAFRSGFSKIDIVLPDPKSKNELTAPLLKLTVGLACAFILVPSADVAIDALTKC 1070
DB 1011 RPLSAFRSGFSKIDIVLPDPKSKNELTAPLLKLTVGLACAFILVPSADVAIDALTKC 1070
QY 1071 ACLLGCRLEIAKCIANPACANVACLOTNNRPDETCQIKGDLFEENSVDENFCAY 1130
DB 1071 ACLLGCRLEIAKCIANPACANVACLOTNNRPDETCQIKGDLFEENSVDENFCAY 1130
QY 1131 SRKRCVPRKSDVGEFPVDRNAVONFNMKDPSGKWTITSGINPTFDA 1190
DB 1131 SRKRCVPRKSDVGEFPVDRNAVONFNMKDPSGKWTITSGINPTFDA 1190
QY 1191 DNLKLVNLSMRITLDSGEFFTRSAVQTFVQDDPLPGALYNHNEFLHYQ 1250
DB 1191 DNLKLVNLSMRITLDSGEFFTRSAVQTFVQDDPLPGALYNHNEFLHYQ 1250
QY 1251 NKPEDYIFYYKGRNDAMGYGSVLYTRSAVLPSSTIPELEKAKSIGRDEFSTFRTDN 1310
DB 1251 NKPEDYIFYYKGRNDAMGYGSVLYTRSAVLPSSTIPELEKAKSIGRDEFSTFRTDN 1310
QY 1311 TCGPEPALVERLEKKEVEGERTIIEVEEIEEVEKVRDK 1370
DB 1311 TCGPEPALVERLEKKEVEGERTIIEVEEIEEVEKVRDK 1370

QY 1371 ODEENFVELSKEMEFLEDEIKMESEVKELEGRALPIRKVR 1412
 DB 421 ODEENFVELSKEMEFLEDEIKMESEVKELEGRALPIRKVR 462
 RESULT 4
 Q26223 PRELIMINARY; PRT: 2269 AA.
 AC 026223:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE RHOPTRY PROTEIN
 OS Plasmodium berghei yoellii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YM:
 RX MEDLINE: 95021522.
 RA KEEN J., STINHA K., BROWN K., HOLDER A.:
 RT A gene coding for a high-molecular mass rhoptry protein of
 RT Plasmodium yoellii.
 RL Mol. Biochem. Parasitol. 65:171-177(1994).
 DR EMBL: L27838; AAA21304.1;
 SQ SEQUENCE 2269 AA: 265158 MW: E0A79FA6 CRC32:
 Query Match 2.6%; Score 194.5; DB 5; Length 2269;
 Best Local Similarity 18.4%; Pred. No. 0.0069;
 Matches 255; Conservative 213; Mismatches 436; Indels 483; Gaps 65;
 QY 304 YILSSQLENNKPDYIFYYGRNDAMDVGSGVYTRSS-----PTLPESTIPNLO--KA 355
 DB 34 YLISNOIKNNLN--VSTYPSGR---EGFTSSLELAKSMETKLEITLITLTKSNEIYVL 87
 QY 356 AKSVGRDNFNTDNGSGPEPLVERLEKTAEBEKL-----LIKEAVEI 401
 DB 88 EKEITELFKKY-----LDEAEKRYLEGLLELNNKIKIOLIAKIEYKNATVEL 135
 QY 402 EEEVEK-----EVEKVRDTEMTLFORLEGEFRELQO--DEENFVELSKEXK 446
 DB 136 KKEIEENNAVIDELANOSPYKTGIEKNKIYNTIKSYEQIYEGIDIFVELSSIVK 195
 QY 447 E-LINLEOMATEVEKLEGRALPI--RKLMAALAPHSNLANHET-----IKYYV 493
 DB 196 EDPIIDIE-DKIKLEMLRSKIDNVYDKIQKMEIETVSKSHLNNIETNNKLPITLIEIKKI 254
 QY 494 GSKLPGRFSGWEDYFGSIYVAKICSSRI--PRYFRKSPRICGLDSGLQLFSH-----550
 DB 255 YDEI--SKELNNKMLEDFKNN--EKELSNKISDYDKKREQL--SEYKSKMLEIRNNHNSQ 307
 QY 550 -----GKHLSPAHISINONV-----KNSGCKFPKDV 577
 DB 308 TNDVNTKEEEKKONKYNCSNEMHTIPI--NEDEISKILISEVTKMKDELISKVNTIYDENKKY 367
 QY 578 ALMWKMGOPAKTAIYAIFILSVASKADAVDAKLTCTCLKECRLELAKCISNPACAN 637
 DB 368 KETVSEHSOFTE-----LIDKIKAYSDK-----ELKK-----397
 QY 638 VACLOTGCKNNRP---DETECOIKCGDLFN-----SVDEFNECAVSKKCKVPRKSD-----686
 DB 397 ---CEOSFNDKNSLINETKNSIE--KEYONITLKLVEYIKVCKSTIKESTIKSSQITL 452
 QY 686 -----VDFEYPPPSVLYOKFDMKDFSGKWEITRGITFPAFCQJLHEFTFE- 735
 DB 453 KDLMLNONTKIVETNSIDS-YIEKFE-QILTKG---QIKLENKFTFESSLNNEHANNEL 507
 QY 735 ---NKLVGNTLSWRITPDGFFTSVAVOKEVQPKYIGILYNDNEELLYODWYLLSS 790
 DB 508 IKFSDKANKLGINENEMLYNOFTER--EKTFNDKIKENI--HINEI-----S 552
 QY 791 KVENSPEDYIF-----VYKGRNDAMD-----GY 814

DB 553 KIEIKIHASIVYINSEETEREIGINIESINTKVEKVENVTNKNKIEKIKHDESDQFK 612
 QY 815 GGSVLYTRS-----AV-----LPSIIPLOJTAOKVGDHFTFI 849
 DB 613 EGNIKYIDKIKINDIMAVSOO IDOHINGLDDIOKSSSEYSEKEDINKLEKVNTEI 672
 QY 850 KTDNTG---PEPPLVRLKCK--VEEGERTIKVEELEE--VEKVRDKEYTLFESKL 901
 DB 673 SMDNVEGIRKKQITIVIKDKKNITVEINKLLEISLEKNDTSLEKVKDINISYGNL 732
 QY 902 FEGFKELODEENFVELSKEXM---DVUDGLKMATVEKLEGRALPIRKLMAY--AT 955
 DB 733 GNLFLE-QDEEKKKAENTIKSMAYIDDIDNKKSSOEIETEMDKMDINKMEMALKIS 791
 QY 956 HCFSTPCDRIKFPSSDDIGIRLGTIRKIRNGTFLKLPIDIOSADLRITGGRSSRPLSA 1015
 DB 792 HDDDKKCHDKSK--NHKEISDIYKSSKILQDE-----SRSDINDIKNLQKNVSE 842
 QY 1016 FRSGFS-----KGIFDIYPLPSKNELKEL-----TAP 1042
 DB 843 SQHNNSDINOCLNEVANINIKL---NKIKIDKVEYSELEKKNKINDELNNSEK 899
 QY 1043 LLKLIVGLACALIVPSADAVDAKLTACALIKGRIELACIAN---PACANVACLOT 1099
 DB 900 VIKRIEGDLS-----LKECRSKINSTLDDKIDICTKNNIVLAK 938
 QY 1100 CNRPDETEC--QIKCGDLFENSVDENEC-----1129
 DB 939 -NINLEETINITHFKNAEENKIVLSNPNINEMADNKSQYILEIKKNGTNDHDIKEL 997
 QY 1129 -----AVSKKCVPRKSDIGEPFAPDPVLYQNFNISDPNGKWTYISGLNPT 1175
 DB 998 KSKHDKSGICYTADQKKALQKKNKLEFQYKEEVVLLNNKYAAVELKNN-----1048
 QY 1176 FDFPCOLHEFTTEGDKNLVGNISWRK-----TIDSGFTFSVAVOKFVODPNQGVLY 1229
 DB 1048 ---FD-----KTKNDSK--QIKKEIKDANHCTLES-----KSEKKYNE-----1083
 QY 1230 NHDNEVLHODQMYLSSKIEKNRPEDYIFYYGRNDAMDVGCGAVYTRSSVLP-----1285
 DB 1083 -IKNEKIHEDS--VANNDSKN-----ATSIKVSVEPKTKI 1118
 QY 1285 ---NSI-----IPLEKAASIGRSTFIRDTNCGPEPALVERIEKTEVEGER 1331
 DB 1119 IKINEIKTKSDQDLKELTNDLEKOISLSDIOETKLTEN--GKLKLTLEELLSLKROK- 1176
 QY 1332 IYKAVEIEIEVEKVEKVGRTMTLFOR-----LAEFNEKODEENFV---RELS 1381
 DB 1176 ---KNIDQKKLEDEVNSKIKNTENTVNOHKKRYEIGIVERIKETAKTNKQIESTELI 1232
 QY 1382 KEEMEFL 1386
 DB 1233 KPTIOHI 1234
 RESULT 5
 Q26216 PRELIMINARY; PRT: 2401 AA.
 AC 026216:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE RHOPTRY PROTEIN (FRAGMENT)
 OS Plasmodium berghei yoellii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YM:
 RX MEDLINE: 97077455.
 RA STINHA K.A., KEEN J.K., OGUN S.A., HOLDER A.A.:
 RT "Comparison of two members of a multigene family coding for
 RT high-molecular mass rhoptry proteins of Plasmodium yoellii";
 RL Mol. Biochem. Parasitol. 76:329-332(1996).

DR EMBL: U36927; AAB41263.1;
 FT NON-TER
 SQ SEQUENCE 2401 AA; 281980 MW; 39440760 CRC32;

Query Match
 Best Local Similarity 2.2%; Score 167.5; DB 5; Length 2401;
 Matches 260; Conservative 274; Mismatches 498; Indels 525; Gaps 71;

QY 23 PCNERFHSRGGPPTIIMK--IRSNNGYNSFLEFSTYSSSDSHCKRQSCSID 80
 DB 453 PNEKMYO--KPSIEIKMDEFLSKYNNKYNDD--KVEKEVESEHNFTLKKITE 508
 QY 81 TSFEIORDLRGNTLLERKQROFOLAIVLCTEFLVPRDAVDAKTAACLECR 140
 DB 509 VSDEIRKYEKKFNDSKSLINETKSTIEEYQINLT--LKKVD--DYIKVC----- 556
 QY 141 IELAKCIANPSCAANVACLOTCNNR--PDETCOIC-----GDLFENSVD 185
 DB 556 LNTNELITN-----CHNKQTTLKADNONIKIKETNSIDKITDKFENLTD 603
 QY 186 Q-----FNECAVSRKCV--PRKSDVGEFPPVDRNAVYONEMKRFSGKMY 229
 DB 604 KTELETKFTGLSLNHNESNNKELLTFYDIKANLKG--KKEMALIKOEKKEKAVE-- 659
 QY 230 ITSGINPTDAPDCOLHEFHENDKLVGLTMRTITLDGFFTRSAVOTEVODDLPGL 289
 DB 659 -----DIKKKNVDINKIVSNITETTYT-----SI 682
 QY 290 YNHND-----EFHYODDYILS-----SOLEKRPDDYIFYYRGRAND 330
 DB 683 YNINDETEIEKSTELLNTR--VLEKYAVANTYNINELKELKDYODF----- 731
 QY 331 GGGSVYITRSTPLPESILPMLKAASVGDENNFTTTNSGCPPLVERLEKTAEG 390
 DB 731 GKENIKRYPDENKIKINDITLNOIKDSIETLEFKKNSNHNDELKQIDKLRV--PN 788
 QY 391 EKLIRKAVEIEEVEKEVEKVRDTEMTLFO--RELEGEFKQDEEN----- 437
 DB 789 KTMENEDPKETIEKINIVERT--DKKNKIKYEDIKLNEISLNDKSLKELKINUSY 847
 QY 437 -----FVELSEKEKEI-----LNELOMATVEKLFGRALPI-----RK 471
 DB 848 GKSIGNLFTLOIDEKKKAETHIKAMEAYIDDLNINKKSOIEKEMININDIKMDIHE 907
 QY 472 LRMAAPHSNLANHEIKYYSKLPGRKRS-----GWEDYFGSI-----VYAK 518
 DB 908 MKALINSHDOYKIYHTTSKNH--EKISIDIRKNSLTIODFSEESYINDIKLEKNVLS 966
 QY 519 ICSRSRIPRFKSPRICGGLDSRGLQLFSGHKLMSAHSINONVKGSGCKFRDVA 578
 DB 967 QNNNDINDIYLSKINI-----YNIILKNIKIKITDK----- 999
 QY 579 LMYWKGQFAKTAIYAFILSVASKADAVDALTCCL-----LKECRLELAKCISN-- 632
 DB 999 --VKEYTDEIEKN-----KKIKNELSSEKIIITOLKENSLEKCKSIKSTIDNAY 1048
 QY 632 -PACAAVACIOTC-----NNRPDETCOIKGDLFENSVDENFCAVSRKCVPRSD 685
 DB 1049 VSECKATNTLKTIVNEKN-----INTYFKN--EEDNO----- 1083
 QY 686 VGDFEVPDPVLYOKFDMKDFSGWFTIRGLNPTDAPDCOLHEFTEENKLVGLMSRI 745
 DB 1083 -----NVLNENINEMADTKSOYTL-----NIKKNGTNTTYNI 1117
 QY 746 RTPDGFTSAVOKEVODPKYPIGLYNHNDNEVLYLSDOWWYLLSKVNSDEYIFVYK 805
 DB 1118 KE-----LKEHKRSKNYKDE-----AGKNTOE-----IK 1142
 QY 806 GRNDAMDYGGSVLYRSLVPESSIIPELOTAQVSGDFETFKTNTGPEPEPLERL 865
 DB 1143 KNKEFEKEQEV-----YVLKRYAVEIK-----NKFKTKN-----YSEOI 1181

QY 866 EKKVEGERTIKVEEIEEVEKVRDKEVTLFSKLFEGRELOREDFENLRLSEEND 925
 DB 1182 IKELKDHNTFTSOADSKSEKKNKEIKNEQIRI-----EEVAKNNKSNKA 1226
 QY 926 VLD-GKMBATEVEELFERALPIRIRKMAVATHC--TSPCHRIREFSSDDIGRLGTR 982
 DB 1227 IIDIOISVEPRKI-----KFLIKDLRTKSDDLKREKDIETKISMISIDTOETKL--IEN 1280
 QY 983 KRINCFPLIKLIPPIQSOILRTTGGRSSRPLSAFRSGFSKGIDIVPLPSKNEKELTAP 1042
 DB 1281 KILWLT--LEKLLESKNO-----KKNIED-----OKKEDEVANSK 1314
 QY 1043 L-----LKLIVGLACAFLLVPSADAVDAKTAACLLKGCRIELAKCI 1085
 DB 1315 IKNIESVNOHKRKYELIGYIEKINEIACA-----NKDOIESYOKLI-----IPTIKNL 1362
 QY 1086 ANPACAAVACIOTCANNRP--ETECO-----IKCGDL-----FNSV 1121
 DB 1363 ISPEKANDLGITFNKLNKRYNTENNNIYEERFKSYDLITHYLETSKEPTTYEQINKR 1422
 QY 1122 VDEFNCAVSRKCVPRKSDLGEPAPDPVLYONF--NISDEGKN--YITSGINPTP 1176
 DB 1423 ITAQNELITNKKVKNKAKSYLDDIANEFDRIVHFKKLNVDVNDKFTNEY--SKVNGK 1480
 QY 1177 DAFDCOLHEFHTEGONKLVGNISMRITLDSGFFTRSAVOKFODPQGVLYNHNDNEYL 1236
 DB 1481 DNISNSINNKKSTENLNLNLOTKEM--YANIVSKKY-----STK 1522
 QY 1237 HYODMYILSSKIEKRPEDYIFYYRGRANDMDYGGAVYVYTSVLPNSIPELEKAR 1296
 DB 1523 YEKENFINIRPLAN-----SLNIOIKSSGIDLFKNIN--ALIPYLDISOCK 1568
 QY 1297 SIGRDEFTFIRDTMCGEPALVERIKTYEBSERT--YKREVEIEEVEKE 1347
 DB 1569 -----DLTFI-----PSP-----EKISEYTKISDSYNTLLDLKRSOLOKKEQA 1611
 QY 1348 VEKVRTMTLPLORLAGFENLKOEDENFVELSKREMFLEIEM--EASEVEKL 1401
 DB 1612 LMLI--FENRLHDKYQATNELK-----DLSLDKKKKEDILNKVLLKHSKNEMLKL 1662

RESULT 6
 Q49525 ID 049525 PRELIMINARY: PRT: 1365 AA.
 AC 049525;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 GN LMP1.
 OS Mycoplasma hominis.
 OC Bacteria: Firmicutes: Bacillus/Clostridium group; Mollicutes:
 OC Mycoplasmataceae; Mycoplasma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MH56;
 RX MEDLINE: 95369882.
 RA JENSEN L.T., LADEFOGED S., BIRKELUND S., CHRISTIANSEN G.;
 RT "Selection of Mycoplasma hominis Pe21 deletion mutants by cultivation
 in the presence of monoclonal antibody 552.";
 RL Infect. Immun. 63:3336-3347(1995).
 DR EMBL: U21962; AAB41013.1;
 SQ SEQUENCE 1365 AA; 154982 MW; 23BE4182 CRC32;

Query Match
 Best Local Similarity 2.1%; Score 156; DB 2; Length 1365;
 Matches 209; Conservative 187; Mismatches 420; Indels 430; Gaps 52;

QY 272 TRSAVOTFVODDPLPGALYNHNDNEFLYODDWMYLLSSQIEKRPDYIFYYRGRANDMDG 331
 DB 282 TRNOIOEFINT-----NKNNP--NYSE-----LSISOLSKRD-----SKNSVTDS 319

QY 332 YGGSVIYTRSPLEPIT-PNLOKA-----AKSVGDENNFTITDNGCPPEPLVERLEK 385
 DB 320 SNKSDIESANTELOKALANADKVOADNLAKSKEOLNVSANNT-----LSA 369
 QY 386 TAEGBELLKEAVEEEVEKEVEKVRDTEMTLFORLLEGF-----KELODEENFVREL 441
 DB 370 KLTDKNTIOOAKTELEKEVQKADQAKSNNTASMOQAKSLDAKVAEITKLETF-----426
 QY 442 SKEKEILLNLOEATEVEKLEFGRALPIRKLMLALPHNPLANETIKYVSGSLGKH 501
 DB 426 NKDEAKFENELKOTRNOIOEF-----447
 QY 502 RFSMGWEDYESTIVAKICSSRRIPIRFRKSPRICCGLDRLGLOFSGKHNLSPAHS-I 560
 DB 447 -----INTKNNPNYSSELISOLTSKROSKSVSDSKSSKDIESANTEL 489
 QY 561 NONVPKSGCGKFPKDYALWMEKMGOFATAIYALFLLSVASKADAVALKTCICLKE 620
 DB 490 KOALAKAN-----ADKVOA-----DNLAKS 509
 QY 621 CRLEIAKISNPACANVACLOTCNNRPDEEOIKGDLFENSVDENECASRKCV 680
 DB 510 IKEQLNNSVSN-----ANLSAKLIDKON--TIQOAKTE 541
 QY 681 PKRSDVGPVPDPSVLVQKFDKMDFGKWEITRGLNPTFDA-----FDQLEHFTHEEK 736
 DB 542 LEKE-----YQKADQAKISNNTASMOQAKSLDAKVAEITKLETFNKDEA 588
 QY 737 LVGNLSWRIRTPDGEFFTRSAVQKVDOPKPYGLIYHNDENYLLYODWYILSKYENSP 796
 DB 589 KENELKO-----TRNOIOEFINTNK-----NNPN-----YSLISOLTSKDS-- 627
 QY 797 EDYIFYYKGRDAMDVGSGVLYTRSAVLPEIIP-----ELQTAQKVGDEFTFK 850
 DB 627 -----KNSYDSSNKSDIESANTELKALAKANADKVOADNLAKSKEOLNVS 676
 QY 851 TDNTGCPPEPLVERLEKVEEGERITIKVEEIEVEVK-----VROKEVTLF 898
 DB 677 NANT-----LSAKLIDKNTIOOAKTELEKEIKANOAKISNNTASMOQAKSLD 726
 QY 899 SKLEFEGKELO-----RDEENFLRELSEEMDVLDGLKMEATEVEKLEFGRALPIRKLMAVAT 935
 DB 727 AKVAEITKLETKNDEKAEKFNELKOTRNOI-----QEFINTNK-----766
 QY 956 HCFSPCHDRIRFEFSSDDIGRLGIRKRINGTFLIKLIPLOQSD-----LRTGGS 1009
 DB 766 ---NNPNYSEL-----ISOLTSKRSKNSVSDSSNKSDIESANTELKQALNTAKK 814
 QY 1010 S-----RPLSAFSGSKGIFDIPL-----PSKNEP-KELTAPLLKLVG 1049
 DB 815 SSIDNELRPL-----KNDIOSKIEEGPIRNTNFSNISKLETKNKLAELT-----863
 QY 1050 VLAACFLIVPSADAV-----DAKTCALIKGRIELACIANPACANVACLOTCN 1101
 DB 863 -----KADAIKNPSSKOALDSQOYOKLGNELKITE-----EFGVEYEN 907
 QY 1102 NRPDEEOIKGDLFENSVDENECASRKCVPKRSDLGFEFPAPD-----PSVLVO 1155
 DB 908 SNIGYRLFKLAQAEQFNNSVDKLNAMBEKOTILSKKQKLGOSKIDVLOLSTEMSRQ 967
 QY 1156 NENISD--FNGKWTYTSGLNPTFADCOLEHFTEGDKLVGNISWRIKITLDSGFTFR- 1213
 DB 968 ESTIKKIVINIOAHIRNLN-----SQYRLEAD-KLIAN-----MKRGYGDV 1009
 QY 1213 --SAVOKEVQDPNPGVLYLNDNEYLHQQDMM-----YIISKTEKPEPIFYVYRGHN 1265
 DB 1010 GISLQKQ--QDLMDSDVSVSDS-----LKDDFKALRYLVODYTKNPSVSMINRMS 1064
 QY 1266 -DAMDGYGAVVYTRSSVLPNSI-----IPELEKAASIGR-----1301
 DB 1065 IENYQUNRLIILVRENEILLDKAKDLKRAEKTIKVDENINSIDORAKRKLKEIILNAN 1124
 QY 1301 DFSTFIR-----TDNTGCPPEPLVERLEKIVEEGENIYKE-----VEEIEVEKEV 1348

DB 1125 DLSNFTLNHOKNOFTAKDITPKISLENNKLNINOYLLPIIKKAVASKISIEKN-KREL 1183
 QY 1349 EKVGRTEMTLFCRLAEGFENELKODEENFVRELSEMEFLDEIKME 1394
 DB 1184 EDIIRSNYLMF-----KVEINKRYISELTNRKOVELSRNINE 1220
 RESULT 7
 Q90337 PRELIMINARY; PRT: 1933 AA.
 AC Q90337;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, last annotation update)
 DE MYOSIN HEAVY CHAIN
 OS Cyprinus carpio (Common carp)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Cyprininae; Cyprinidae; Cyprinus.
 RN [1]
 RP SEQUENCE FROM N. A.
 RC TISSUE=FAST SKELETAL MUSCLE;
 RX MEDLINE: 97352533.
 RA HIRAYAMA Y., WATABE S.;
 RT "Structural differences in the crossbridge head of
 RT temperature-associated myosin subfragment-1 isoforms from carp fast
 RT skeletal muscle".
 RL Eur. J. Biochem. 246:380-387(1997).
 RN [2]
 RP SEQUENCE OF 738-1933 FROM N. A.
 RC TISSUE=FAST SKELETAL MUSCLE;
 RX MEDLINE: 97176447.
 RA IWAI J., HIRAYAMA Y., KIKUCHI K., KAKINUMA M., WATABE S.;
 RT "cDNA cloning of myosin heavy chain isoforms from carp fast skeletal
 RT muscle and their gene expression associated with temperature
 RT acclimation".
 RL J. Exp. Biol. 200:27-34(1997).
 DR EMBL: D89990; BAA22067.1;
 DR EMBL: D50474; BAA09067.1;
 DR PFM: PFM0612; IO: 1.
 DR PFM: PFM0063; myosin_head; 1.
 KM myosin.
 SQ SEQUENCE 1933 AA; 221092 MW; 43DE2593 CRC32;
 Query Match 2.18; Score 156; DB 13; Length 1933;
 Best Local Similarity 17.0%; Pred. No. 0.9; Indels 356; Gaps 42;
 Matches 203; Conservative 181; Mismatches 454;
 QY 379 LVERLEKAEGERKLLKEAVEIEEVEKEVEKVRDTEMTLFORLLEGFKE-----L 430
 DB 606 VVOIYKSKALVYLAIVYVPAEPAAGKKGKGGK-----EQYSAVREMLGLMTNL 661
 QY 431 OODEENFVRELKEKE-----ILNELQNE-ATEVEKLEFGRALPIR-----471
 DB 662 RSTHHPVRCILIPNSKTPGLMENFLVHOLRNGVYEGIRITKFPSPRIHYGDKQRY 721
 QY 471 -KIRNALAPHSNFLNHTIKYVYGSKLPGHKRFSMGWEDYF---GSIYVAKICSSRIP 526
 DB 722 KYLNASVPEGGFTDNKKAETKLGSDIDVDHNOYKRECHITVFFKAGLLGLLEMRDEKLS 781
 QY 527 RYFRKSPRICGLDRLGLOFSGKHNLSPAHSINONVPGNSGCKFPKDYALWMEKMG 586
 DB 782 HLYTMQALARGYVNR--KEFYKMERREALYSIOYIRSFMYKHP--WKVYFKIK 836
 QY 587 OFAKTAIYALFLLSVASKADAVDAKTCICLKECPLEIAKICSNPACANVACLOTCNN 646
 DB 837 PLKTAESE--KEMASKENFEKMKEDLTALAKKLELEKMYSLVQEKNDLLQVTSF 893
 QY 647 RPDTEEOIKGDLFENSVDENECASRKCV-----VDEFNECANSRK-----CVPKRSY 886
 DB 894 SENISDAEBCSLISKIQLGKLEKTEETRLDEDEEELINAEITAKKRLLEDECSSEIKDI 953

```

QY 687 GDFPVPDSVLVQKDKMDFSGKWEITRGINTPTDAFDCOLHEFTENKLVNLSWRIR 746
DB 954 DDLLETLAKVEREK-----HATENR-KVNITEEMA 982
QY 747 TPDGFFFRSVQKFFVODPYKGLYNHNDWYLLSYDDWYLLSKVENSPEYITVYVKG 806
DB 983 SOD-----ESIAKLTREKALQAFHAGQTLDDQAFEDKVNLTAKIKLEQV----- 1031
QY 807 RNDAMDYGSGSVLITRS-----AVLPSTIP--ELQTAQKVR----- 844
DB 1031 -----DDESGSLDEKRLRNDLDRVYKRLLEGDLKLAQESIMDLENKQSDSEKIKKDFE 1085
QY 844 --DNNTIKITDNTGPE-----PILVERLEK-----KVEGERTIIEVEI 883
DB 1086 ISQFLSKIDEOSIGAQLOKIKELQARLEELIEEELERAKRKAQVQNRADLSRELEI 1145
QY 884 EEEVER-----VRDKEVTLFSLFEG 904
DB 1146 SERLEEGGATAQIEMNKKRAEFQKMRDLEESTLQHEATTAALRRKQDSVALEGEQ 1205
QY 905 FKELQDENEFLELSKEEDVLDGLKMEATEVERKLPALPIKLAVAVTHPTSPCHD 964
DB 1206 IDMIQRIKQKLEKSEKSEKMEI--DLSNMENAVAKAKN--LEKM-----CRT--LBD 1253
QY 965 RIRFF--SSDDGIGRLGTRKRIKNTFLIKLIPQISADLRTTGRSSR-----P 1012
DB 1254 QLSEIRAKSDENSROLNDMN-----AQRARLQTENGEFSRQLEKALVSG 1299
QY 1013 LSAFSGSEKGIIDIVPLPSKNEKELTAPLLKLVGLVACAFIYPSADAVALKTCAC 1072
DB 1300 LTRGKAFTQILEDL-----KRHVE----- 1321
QY 1073 LKGRGRIELAKIANPACAAVAC--LQTCNNRPDETECOING-----DLE 1118
DB 1321 -----EVAKAKNALAAVOSARHDDLLREKYEEEOEAKALQSGMSKANSEVAOMKAYE 1375
QY 1119 NSVDEFNECAVSRKCCVPRKSDGEPFAPDPSVLYONFNISFNKKTITGCLNPTDA 1178
DB 1376 TDAIQRTELEBSKKLAQRQDAE-----SIFAVSK--CASLEKTQR 1419
QY 1179 FDCQHEHTEGD--NKLGNISWRIKTLDGFFTRSAVQKFEVDPNPGVLYNHDNYL 1236
DB 1420 LQSEVEDLMDGERANLANLDRKQRPD-----KYLADMOKK--YEESQAL 1466
QY 1237 H-YDDWYLLSKIKENPEDYIFVYRGRNDAMDYGAVVTRSSVLPNSTIPELEKAA 1295
DB 1467 EAAKFAKSTELFKKKNKY-----EALDR-----LELIKREN 1501
QY 1296 KSIQRDSTIRTDNCGPEPALVERIEKTEVEGERIIVKEVEIEEVEKEVEKVGRT 1355
DB 1502 KNLQOELISDSEGLGETGSIHEIEKAKTVESEKAEIQTALHEKGLTLEHEBSKILRVQ 1561
QY 1356 MTLFQ-----BLAEGNELKODENFVRELKSEMEFLDEIKV--EASVER 1400
DB 1562 LEINQVSEIDRLAKDEMEQIKRNSORVLDSSQSTLDSVSRNDALRVKK 1615

```

```

RX MEDLINE: 98049343.
RA KLEKH H. P., CLAYTON R. A., TOMB J. F., WHITE O., NELSON K. E.,
RA KETCHUM K. A., DODSON R. J., GWINN M., HICKEY E. K., PETERSON J. D.,
RA RICHARDSON D. L., KERLAVAGE A. R., GRAHAM D. E., KIRKNESS E. F., DOUGHERTY B. A., MCKENNEY K., ADAMS M. D., LOFTUS B.,
RA KIRKNESS E. F., DOUGHERTY B. A., MCKENNEY K., ADAMS M. D., LOFTUS B.,
RA OVERBERK R., GOCAYNE J. D., WEIDMAN J. F., McDONALD L., UTTERBACK T.,
RA COTTON M. D., SPRIGGS T., ARTIACH P., KATNE B. P., SYKES S. M.,
RA SAOON P. W., D'ANDREA K. P., BOWMAN C., FUJII C., GARLAND S. A.,
RA MASON T. M., OLSEN G. J., FRASER C. M., SMITH H. O., WOBESE C. R.,
RA VENTER J. C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-376(1997).
DR EMBL: AE001032; AAB90211.1;
DR TIGR: AF1032;
DR PFAM: PF00470; RecF; 1.
KW Hypothetical protein.
SQ SEQUENCE 886 AA: 103633 MW: 865847 CRC32:

```

```

Query Match
Best Local Similarity 2.1%, Score 155.5; DB 1; Length 886;
Matches 137; Conservative 85; Mismatches 221; Indels 221; Gaps 27;
QY 797 EDYIFVYKGRNDAMDYGGSV--LYTRSAVLPESITPELQTAQVAGDNTFTKTD-N 853
DB 152 EDY-----ENAKNKGAVIRMLERERKEKLEFFSOEIOIKRQKEK-----KAEIE 197
QY 854 TCGPEPPLVERLEKVEGEERTI--IKVEEIEEVEKVRDEKVLFSKLEPGEKLEOR 910
DB 198 RSEIKRSIESLREKLSSEEVANLESRLKELEKRSLESIRQESSVLDV-----R 249
QY 911 DEENFLRELSKEEDV--LDGLKMEATEVEKLFQALPIKRLMAVAHCTSPCHDIR 967
DB 250 GLEKRLREQLQLEVEYRIEDLEKRAKVEKEL-----KPRAEY- 290
QY 966 PFSSDDGIGRLGTRKRIKNTFLIKLIPQISA--DLRTTGRSSRPLSARSGSKIF 1025
DB 290 -----SILEKLSEINQALNRVERREDLREAGIQAOLKRAE 329
QY 1026 DIVPLPSKNEKELTAPL-----LKLGVLAACAFIYPSADAVALKTC 1070
DB 330 D-----NSKIEBITKRIEELERELERPEKSHRLE-----TLKPKMDRMQIK- 373
QY 1071 ACLIKGRIELACIANPACAAVACIQTNNRPDETECOIKGDLFENSVDVDFNCAV 1130
DB 373 -----AKLEKNLTPDKVE--KMTDLSKAKEE-KEITE 404
QY 1131 SRKCVPRKSDGEPFAPDPSVLYONFNISDFNGKWYITGSLNPTDAFDCOLHEFTTEG 1190
DB 405 KKLKLIKKSLSLTRAQILKAAVE-----LKSARETCVPCGREIDEBERKN 451
QY 1191 -----DNKLGNISWRIKTLDGFFTRSAVQKFEVDPNPGVLYN-- 1231
DB 452 IMAEYTRMKRIAEELAKADETEKKLERLEKVEKALEKQETVYKTRQMWDELKALENEL 511
QY 1231 --HDNEXLHYDDWYLLSKIKENPEDYIFVYRGRNDAMDYGAVVTRSSVLPNSTI 1288
DB 512 SSHDAE-----KLSAESE-----YKRVKREDGLRGQOKILLSA---SRI 550
QY 1289 PELEKAASIGDFSTIRTDNCGPEPALVERIEKTEVEGERIIVKEVEIEEVEKEV 1348
DB 551 KELKSLREI-----EAL-KVNESRGGLHKKIREPEGESLEELEREV 593
QY 1349 EKVGRTEMLQRLAEGN--ELKODENFVRELKSEMEFLDEIKVASEVEKLFQKA 1405
DB 594 -----QSLRPFLYKWLKLDASRLSELELKRRE-KLEDEISEIAKLEANGKA 641
QY 1406 LPIR 1409
DB 642 EEIR 645

```

Thu Oct 14 07:57:26 1999

us-09-075-375-2.rspt

Page 7

RESULT 9
ID 025662 PRELIMINARY: PRT: 1939 AA.
AC 025662:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1996 (Tremblrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium cibaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-96V:
RA WERNER E.B., TAYLOR W.R., HOLDER A.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U43145; AAC63403.1;
SQ SEQUENCE 1939 AA; 22901 MW; 5FC4E4C2 CRC32;

Query Match 2.0%; Score 151; DB 5; Length 1939;
Best Local Similarity 17.5%; Pred. No. 1.7; Indels 422; Gaps 50;
Matches 189; Conservative 167; Mismatches 299;

QY 380 VERLEKTAEGEKLKEAVEIEEVEKEVEKVTOTEMTLFQRLLEGFKELQD--EEN 436
DB 95 VKEVEVIEEELKEL--KLEKEQINKKEKELNKSFEIKROMELKELNINIKEN 151
QY 437 FVRE--LSKEKEILNLOME--ATEVEKLFGRALPIRLKLMALAPNSFLANHE 487
DB 132 KINKEIITLKEEK--LNDIESEYIEKNEKEKL--NEYVINLMSL-----D 196
QY 488 TIKYVSGKLPGHKRFSGWMDYFGSVVAKICSSRIPIYFKSPRICGDSGLQLF 547
DB 197 KLTCEVOEKKNLEKIN-----KVIEKENLRLEKFEKNEKIELIESLDG----- 243
QY 548 SHGKHNLSPAHSINONVFNKSGCKFPKDVAMWEMWGFATVAIFILSVASKADA 607
DB 243 -----TIN-----DKKNVEKELT-----SFEKRM 264
QY 608 VVALKTCTCLKECELELAKCISNPACANVACLOTGNNPDETECOIKGDLFENSVD 667
DB 265 IEMLD-----SKLEKEENPAN-----KOAKLE 287
QY 668 ENECAVSRKCYPRKSDVGPDPVSVLYOKFMDKDFSGKMTTGLNPTDPAFCOL 727
DB 288 KENIIEIKLKDIESRE-----KDFRSK-----EKFASMENEL 321
QY 728 HEPTTEENKLVNLS--WRITPDGFFTRSAVOKFVODPKYFGILYHNDNEYLLYODWY 786
DB 322 NTLKSDLSKNACOMEYKLEIKD--LSOSLYEKKE-----IFEIKNEY-----DD-- 366
QY 787 ILSKVENSPEDYIFVYKGRNDAMDYGGSVLYTRSAVLPSITIPLOJIAOAKVGRDN 846
DB 366 -----KINMKELSSI-----ND--KGIDNTVLHSEEXT--NKLKEKTELNEJHKKYN 413
QY 847 TFIKTNTGPEPLVERLEKVEGESERTIKVEEIEEVEKVR-----DKEVT 896
DB 414 LEIE-----TIKNLEKEKELEENKKAHTVEVNTNLKEIK 449
QY 897 LFSKLFEGFEKELORDE--ENFLRELKSEEMDVLGLMEATEVEKLFGRALPIRKMLA 952
DB 450 LLEKTEDEAKHEKHELMELNNOJSLINKER--DNINKMENEL----- 491
QY 953 VATHCTSPCHDRIFESSDDOIGRLGTRKRINGTFLKLILPPIOSADLRTTGRSSRP 1012
DB 491 -----NDKISSLSE-----VN-----ILKKDKOTIG----- 513
QY 1013 LSARSGSGKGFIDVILPJSKNELKELTAPLLKLVGLACAFILVSAADVAKLTCAC 1072
DB 513 -----NDIKTLN-----DLINKLN----- 528

QY 1073 LKCGRIELAKCINPACANVACLOTGNNRPE-----TECOIKGDLFENSVD 1123
DB 528 -----EINTSDKNKMKEDDLAMNEEKGK-----VVID 558
QY 1124 EF-----NECAVSRKCYPRK--SDLGEFPAPDPVSVLVONFNISDPNGKWTITSGINLTF 1176
DB 559 EIEKKYKNEIFLEKEKELKENYADLND-----EISLRNSIYKE--KEFI----- 604
QY 1177 DAFDCOLHEFTEGCKNLVGNISWRITLDSGFTRSAVOK--FYODPNOPGVLYHNDN 1233
DB 604 -----EMKEFTY--ENKI-----NLFNKFEKKNIYEELNLSLTKYDNO 642
QY 1234 EYLHYDDVYIT:SKIKENPEDYIFVYGRNDAMDYGGAIVYTRSSVLPNSIIPELK 1293
DB 643 GLIKQIDELNIOKLTKEEK--YLOLY-----NDN-----MHMPSITCKIDMYSN 687
QY 1294 AANSIGRDEST--FI--RTDNTGPEPALVERLEKTEVEGERIYKEVEEIEEVEKEVEK 1350
DB 688 IKGSDLVFVTAVYIKRRDSSSDANPDITTK--EVALEKRRHAIVALEEKKHEKELIAK 745
QY 1351 VGRTEMTLFQRLLEGFKELNE-----LKODENFVRELKSEEMFLDEIKMASVYELFG 1403
DB 746 LGBGHKEVLRIGBQHKHEETIILEEKHKDVYTKLGEOKHKNITIKLEEHKDVYTKLG 802

RESULT 10
ID 013246 PRELIMINARY: PRT: 3113 AA.
AC 013246:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1996 (Tremblrel. 08, Last annotation update)
DE NUCLEAR PHOSPHOPROTEIN MITOSIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95379843.
RA ZHU X., MANCINI N.A., CHANG K.H., LIU C.Y., CHEN C.F., SHAN B.,
RA JONES D., YANG-FENG J.L., LEE W.H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RT that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029(1995).
DR EMBL: U30872; AAA82935.1;
SQ SEQUENCE 3113 AA; 357281 MW; 605779C3 CRC32;

Query Match 2.0%; Score 147; DB 4; Length 3113;
Best Local Similarity 18.4%; Pred. No. 5.6; Indels 434; Gaps 60;
Matches 240; Conservative 188; Mismatches 443;

QY 379 LVERLEKTAEGEKLKEAVEIEEVE-----KEVE--KVDPTEMTLQRLLEGFKELQD 432
DB 946 LSEILSLEKREMSIILNKRREIELLOENGLKEINASINOENKMLQK--SESFANYID 1004
QY 433 DEENFVREL--KEKEIL----- 450
DB 1005 EREKISISLQYQOEKILLDORCETGNAYEDLSOKYKAAOENKSLCEILNCTSLCE 1064
QY 450 -----NELOMEATEVEKLFGRALP--IRKLMAALAPHSNFLANHETIYYGSLPHG 501
DB 1065 NRKNELE--GLKAFKHEQHEFLTCLAFAEENONMLIETVOALRSEKTDQONS 1120
QY 501 KRFSGWMEDEFSIYVAKICSSRIPIYFKSPRI-----CCGLDSRGLOLFSHGK 551
DB 1121 KSPAGGLKOEITLKEEONKMOKEVNDLQEDQDMKVMYTKHQCQLBSEPIR----- 1175
QY 552 HNSPASHINOVAPKNSGCKPKDVAMWEMWEGFAKTAIYVIFLVSASADAVDL 611
DB 1175 -----NSVKEPESERNQ--CNFKPQMDLEVKEIISDSYNQVQLEMLRNKELKQOSE 1227
QY 612 KTCCTL-----LKECRREL--AKCISNP-----ACAAN 637

```

Db 1228 KKKCKLQHETQIRGDLSTSLNODMOSQETSLKCCCEIDAEKYSIGPHELSTQNDNAH 1287
QY 638 VAC-IQTCNNRPETE--COIKCGDLFENSVDDEFNECAVRKKCY--PKR--SDVGD- 690
Db 1288 LCCSLDTYMKKLNKEIKELCEILOAEKYE--LVTELND--SRSCITATKMAEEYKLL 1342
QY 690 ---FVDPDSVLVOKFEDMKDESGKWFITRGILNP-----TFPAFC 725
Db 1343 NEVKILNDSSGLLHGEIVEDIDGGEF--GEQPNQHPVSLAPLDESNEYHLLTSDREV 1399
QY 726 OLHEPHEEENKLVGNLSWRIRTPDGGFETRSAYVKFQODKYYGILYNHNEILLQD- 784
Db 1400 QMHPFELDER-----FLSLQSEHKILHDOHC--MSKMSSELQTYVDSL 1441
QY 784 ---DWTLLSKVNSPEDEIFYYKGRND-----AMDVGSVLYLRSAYL 826
Db 1442 KAENLYLSTNLRNFGDLVAKEMQIGLEGLVPSLSSSCVDPSSSLSDSSFY--RALLL 1499
QY 827 PE---SIIPELQTAAC--KYGRDPNTEIKTDNCGEPPL--VERLEKKEVEGE 873
Db 1500 EOTGDMSLSLNLEGAVSANOCVDEVEFCSSLOEBENITRKETPSAPAKGVEELSLCEVYR 1559
QY 874 RTIIEVEEIEEVEYK---VRDKEVTLFCKIEG-----FKELQDEENFLRELSE 922
Db 1560 QSL---EKLEKMEESGIMKKEIODELBOLLSEROLDCIRKOYLESNQMOKILTV 1615
QY 923 EMDVLDCLEKATEVEKLF-----GRALPIKLMVATH-----CFTS- 961
Db 1616 TLEMSKLAEEKOTOLSLLEVARLOQIGLSSRSLGIDEDALOGNESCDSIKR 1675
QY 961 ---PCHDRIRFSSDQIGRLGTRKRINGFLKILRPI-----OSADL- 1003
Db 1676 HTEETTERPKND-VHQLDCKADQDLNDEIKETITGALK--PTGECGSDSPPTNE 1731
QY 1003 ---RTTGRS-----SRPLSAFRSGFSKGIPIYPLPSK----- 1034
Db 1732 PGECKTQSSSEISSELSFSGPNALVPMDFIGNOEDJHNLOLRKETSNNENLHLVID 1791
QY 1034 ---NELKELTALLKILVGLACAFLLVPSADAVALKTCACILKGRLEIAC 1084
Db 1792 RDRKVESILNKEKELDSKHLHGEVOLMT-----KIEAC-----TELKTI 1830
QY 1085 IANPCANAVACLOTCNNRPDETCQIKCGDLFENSVDDEFNCAVSRKKCYPRKSDLG- 1144
Db 1831 V-----GEL-----KK--ENSDISE 1843
QY 1144 ---EFPADPSVLYQNFNISPNGKWITTSGLNPTFDFACOLHEFHTE--GDKKLIGN 1197
Db 1844 KLEVFSCDHOELLORVETSE-----GLNS--DLEMHADKSSREDIGDNVAKVN 1889
QY 1198 ISMRKILTDSGFTTSAVOKFVODRPNQPGVLYNHNEYLH-----YQDDWITLSSKTEIK 1252
Db 1890 DSMKERELD-----VENELSRIRSEKASTIEHEALTYLEADLEVYQTEKLCERKDNK 1941
QY 1253 P-----EDYIEFYGRDANDGCGAVVYTRSSVLTNPSIIPERKAKSIGDPSTFI 1306
Db 1942 QKVIYCLEEELSVYSEKNOI--RGEIDTMSKRTTALDQSEKMEKETOGEHOSQEC 1998
QY 1307 RUDNTGPPALVERIEKIVGEGERIYKVEEIEEF--VEKVEVVGTEMTLFORLA 1363
Db 1999 HCIQVAAEAVKREKTELQTLSDVSELKDKTHLOEKLOSLQESKQALSTLTKCELENQIA 2058
QY 1364 EGFNELKQOEENFVRE-----ISKEMFLDEIK-NEASEVER 1400
Db 2059 ---QNKKEKELVKESESLQARLSESDYEKLNVSKALFALAYEK 2099

```

```

DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DE 01-NOV-1998 (Tremblrel. 08, last annotation update)
DI NM1P.
GN NM1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972:
RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR SGD: L0001287; NM1.
DR PRAM; PF00169; PH: 1.
SQ SEQUENCE 2748 AA; 313030 MW; FEEETEDD CRC32;

Query Match
Best Local Similarity 1.9%; Score 146; DB 3; Length 2748;
Matches 283; Conservative 232; Mismatches 581; Indels 486; Gaps 72;

QY 62 SFSDSHCKDKNSQICSDITSEEFQRFDLKRGMTLLKEMKRF-----IQLAIVLC 115
Db 748 SAYEDLVCKENPDY-----EFLKESAKLGHITVNSSEYSLQKYSLEKEVEQS 800
QY 116 TFIYPRVAVALATCACLKECRILEAKTIANPSC-----AANVACLOTCNNRPDET 169
Db 801 LALIVHAATQD---HLSDSAVEELVCKENPDVEFLKESAKLGHITVNSNEAYSL 856
QY 170 ECOT-----KCGD--LFENSVDQFNCAVSRKKCYPRKSDVGEPPVDRNAV 215
Db 857 EKLQPSLAYLVEHAKATDHHLLSDSAIEDLVK-----KENSVDY-EE----- 900
QY 216 VQNFNMKDFSGKWITTSGLNPTFDFACQLEHFMENDKLVNLTFRITLGGFTTSA 275
Db 900 ---LKESKAKLGHITVNSNEAYSELEKKLEQ-----PSLAVIYHAKATDHHLLSDSA 948
QY 276 VQTFVQDPLDGCALYNDNEFLHY-----DWTYLLSQIEENKPDYIFVYVYRG 324
Db 949 YE-----DLVCKENPDVEFLKESAKLGHITVNSNEAYSELEKKLEQPSLEYLVE 1062
QY 325 RND-----AMDVGGSV-----IYTRSPPLPSIIPN-----LOKAASVGRDF- 364
Db 1003 TNHLLSDSAIEDLVCKENPDVEFLKESAKLGHITVNSNEAYSELEKKLEQPSLEYLVE 1062
QY 364 ---NNFITDSS-----CGREPLYERLEKTAEBEKKLILEAV-ELIEEVEKE- 409
Db 1063 HAKATNHLLSDSAVEELVCKENPDVEFLKESAKLGHITVNSNEAYSELEKKLEQPSLEYLVE 1122
QY 409 ---VEKYVDTMTLFORLLEGEFKELOODENENFRELSEKEKELINE--LQMEA-TEVEKLF 463
Db 1123 YLVEHAKATNHLLSD--SAVEELVCKENPDVEFLKESAKLGHITVNSNEAYSELEK- 1179
QY 464 GRALPIKRLMALPNSNLANHETIKYVSGKLPGRKRSWGMEDYFGSIYAK----- 519
Db 1179 ---KLEQ---PSLAVIYH-----AKATDHHLLS--DSAYEDLVCKENPDY 1217
QY 519 ---ICSSRRIPRYFRKSPRICGDSRGL-QLESHKHLNLSAHSINOV 564
Db 1218 EFLKESAKLGHITVNSNEAYSELEK-----LEQPSLAVIYHAK-ATDHHLLSDSA 1268
QY 565 PRGNSGCKFPKDVALLMWKMGQFAKTAIATITLVSADK-----AVDALKTCTCL 617
Db 1269 YEDLVCKENPDVEFLK-ESAKLGHITVNSNEAYSELEKKLEQPSLEYLVEHAKATNHLL 1327
QY 618 LKECLE-LANCISNPAC-----AANVACLOTCNNRPDETQIKCGDLFENSVDDEFN 670

```


Db 1328 LSDSAVEDLVCKENPDMFELKESAKLGHVTVSKKEYSLEKLEPSPSL----- 1378

QY 671 ECAVARRKCCVPRK-SDVGD-----PVPPSVLYOKFDMKDSGKWFITTRGLNPTDAF 723

Db 1378 EYLVAHAQIOISKIISDFNTLANPSMEDMASKLOKLEYOISNDEYI--ALKNTMEKP 1435

QY 724 DCQ-----LHEFH-----TEENKLVGNISWRIRPDGFFTRSAVOKEVOD-PKYPGILYN 773

Db 1436 DVELLSKILKGYHIIIDTTYNELVSNFN-----SPLKFIKESAKSKGYRLI 1482

QY 774 HONEYL-----LYODDWTLLSK-----VENSPE-----D 798

Db 1483 EPEEYIDLNRIATTSKEEIDNFCQIGCALDSKEVERLKNLSNPKKFIENNALD 1542

QY 799 YIFV---YKGRNDAMDYGGSVLYTRSAVLPESITPELOQAQKVRDNTFTKNDNTC 855

Db 1543 LVLVDKTEYOAMKD-----NASNKSILPSTKAL-----DVTM----- 1577

QY 856 GPPPLVERLEKKEVEGERTIIEVEEIEEVEKVRDKEVTLFSGLEGEKELOD----- 912

Db 1577 -PAPQLASAKSSLOK-----RILSDIENELKAL-----GYVAIKKEMLPN 1616

QY 912 -EENFLELSKEEM-----DVLGKMEATEVEKLEF-----RALPIR 948

Db 1617 LEKPIYONASKNDVNLNCSKFSVLPSTEEYDNMRKHTKILNTIGDPSIDFLKEKEKY 1676

QY 949 KLAANVATH-----CETSPCHDIRFFSSDDGIGRLITRKIRNGFLKILPPIOSAD 1001

Db 1677 QMLTISHDYEEKOEALENPGYFIELEKASALGY-----ELVESEVLEDRMKQIDSPD 11729

QY 1002 LRTIGRSSRPLSAFRSGFSKGIPIVPLPSKNELKELAPILLKLVGLACAFIIVPSA 1061

Db 1730 IDYQEKAAAR-----EMWYL--RNEEKE--ALOKKIEYPSITFLIEKAA 1770

QY 1062 -----DAVALKTCACILKGCRIELAKCIANPACANVACIQCNNRPDETECOIKCG 1114

Db 1771 GNNKLLVDOIEYDET-----IRK-----NHPTMELESCCH 1802

QY 1115 DEFENSVDENECVARRKCVPRK-SDIGEPAPDPVSVLVONFISDNGKWTISGLN 1173

Db 1803 HL--NLVILLDQNEVSTLEPLENNVEDLINTLSKLYIAPTIVIODLIGKE-----N 1855

QY 1174 PTFDAFCQLEHFEETEGNKLVGNIWRIKTLDSGFTRSAVOKEVODPQSPVLY----- 1230

Db 1856 PNEVDYLDKSLNK-----WDYVAISRODYELMAKAYKPKQLDVLKIS 1896

QY 1230 -----NHUNEYLAHYODDWTLLSKFIENKPEDEYFYVYRGRNDAMD 1270

Db 1897 SEKIDHIVPLSEYNLMATNPNPSLSYLKERAVALNHHILKEDDYKNIL-----AVSE 1950

QY 1271 YGGAVVYTRSSVLPNSIPELEKAKSISGRDPSFIFRDNTEGCEPALVERIEKVEEGE 1330

Db 1951 HPFVILHSEKASILNKVLVDKD-----DPAIKSRSEI--KPIIDILSTIKALSMG- 1998

QY 1331 RIIVKEV-----EIEEVEKEVEKVGRTMTLFORLAEGENELKODEENFVRLSEEM 1385

Db 1998 KILVNESTHRKNKLSLSPDSSEFLIMKAKEOGLIISKEYSELDOIDRPSLDVLKEKA 2057

QY 1386 EFIDEIKMEASEVEKLEFGALP 1407

Db 2058 AIFDSIIVENIEYOOLVNTTSP 2079

RESULT 12

Q26024 PRELIMINARY: PRT: 839 AA.

Q26024: 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)

DE TOPOISOMERASE I

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 96001260.

RA TOSH K.T., KILBEY B.J.K.;

RT "The gene encoding topoisomerase I from the human malaria parasite

RT Plasmodium falciparum";

RL Gene 163:151-154(1995)."

DR EMBL: X83758; CA58716.1; "

DR PFAM: PF01028; Topoisomerase_I. 1.

KW isomerase; Topoisomerase.

SEQUENCE 839 AA: 98109 MW; 1CAF3F0 CRC32.

Query Match 1.9%; Score 144.5; DB 5; Length 839;

Best Local Similarity 17.3%; Pred. No. 1.3;

Matches 191; Conservative 176; Mismatches 390; Indels 347; Gaps 46;

QY 352 LOKAKSVGDFNNFTTNDNSGPPPLVERLEKTAEGEKLKEAVEI--EEVEKEV 409

Db 24 INKIKOMLG--NN--KSCNSRSKRESIKOKSNSLGIKNTKKSIGIKKEEKKKQI 78

QY 410 EKVADTEMTLFORLLEGFPELOODEENFVRELSEKEEILNELQMEATEVEKLEFGALPI 469

Db 79 SKRSNLEKERNKKEKKKYYEKKSRV--KDETKLVNKKKEIÖNNKK-----PK 128

QY 470 RKLMAIAPHSNPLANHEITIKYVSGSKLPGRKFRSGMWEDYFGSLVAKICSSRIIPRYE 529

Db 129 KILKKS--EENF--EPINRW-WEKIDQTDIDQNNY----- 159

QY 530 RKSPPICGLDSRLOLFSGHKNLSPASINONVPRKSGCKGPRDVALMWEXWGOFA 589

Db 159 -----LEHRDL-IF-----SPRY-VQHNV-----IFKSIKIEL--NA 188

QY 590 KTAIVAFILSVASKADAVDALKTCTCLKECPLELAKCISNPACANVACIQCOTNNRPD 649

Db 189 KSEELATWCS-----AIGSDYCKEKKFLINFEKFTFN-----SLENDNIHQ 231

QY 650 ETECOIKGDLFENSVDENECVARRKCVPRKSDVGDPEVPPPSVLYOKFDM----- 704

Db 232 ENETKILKKGIDISNFEID-----FMPIDHLLKREKLNKTEE 271

QY 704 KDSGKWFITRILNPTPDAFCQLEHFEETENKLVGNLSWRIRTPDGG-----FFRSAY 758

Db 272 KEEEKMNRKELPTVYALVMDIREKISSNKAEPGIFRGGEHPKQGLKKRIPEDEV 331

QY 759 QKFVODPKTPIGLYN-----HDNE--YLLYODDWTLLSKVENSPEDEYFYVYK 806

Db 332 INISKDAPVPRIYDNCGHNWDIYHDKVTWLAAYD-----SINDQIKYTFLS 381

QY 807 RNDAMDYGGSVLYTRSAVLPESIIPELOTAOKVGRDNTFIKT-----DNTGPEPPV 862

Db 382 AQSFKCYKRLMYENAR-----KLKSCVHKIREDYKKNKKNINIDKQIGTAVYLLI 433

QY 863 ERLEKVEEGEETIIEVEEIEEVEKAVDKVTLFSLFEGFELQDEENFVRLSEKE 922

Db 434 DFLALRV-GGEK-----DIDEADTV-----GCCSLRREHISFADHPK 472

QY 923 EMDVLDGLKMEATEVEKLEFGALPIRKIMAAVATHOFTS--PCHDRIRFFSSDGTIGRLG 979

Db 473 SVD-----SKEKTNDEKVNKIPLPNLESISSEDCYITLDELFGKDSIRYFT----- 521

QY 980 ITRKRINGTFLKTLPIPIOSADLRTTGSSSRPLSAFRSGFSKGIPIVPLPSKNE-LKE 1038

Db 521 --VKIDKQAVYTAII-----IFCKNKNRDEGVFDDIOTCSKLENYIKE 559

QY 1039 LTAPLLKLVGLVACAFIIVPSADAVDAL--KTCACILKGCRIELAKCIANPACANVAC 1096

Db 560 IMPTLSAKVFTYNASTITDQQLKRIKEYGKTTSLYSG-ELELHK-----SK 607

QY 1097 IOTCNRRPEYFCOIKGDLFENSVDENECVARRK-----KCVPRKSDVGEFAPDPSPV 1152

Db 608 KRKSHLSDFTNILDASDSTINDVNNEDENGINKKLSYATTVGKENVDKNSP-IEV 666

QY 1153 LVONENISDPNGKMYITSGNPTFADFQCLHEFHTEGDKLVGNISWRIKTLDSGEFT 1212
 DB 667 DVSNIN-----ELNFINNANREVALICNQRSIPKQHTM 703
 QY 1213 SAOKVQVDPNPGVLYNHND-NEYLAHODWYILSSKIENKPEYIFYYGRANDAMDGY 1271
 DB 704 SKIKOIE-----LYNEDIKEYKY-----LOHLKNSDKFIY-----739
 QY 1272 GGAVYTRSSVLPNSIIPLEKAKSIGHDSFTIRNTNCGPPEALVERIEKTEVEGER 1331
 DB 739 -----SKYSTLDGLRP-----NKVKENMK 758
 QY 1332 IIVKEEELKEVEKEVEKVGTEWTEFLQRLAEGFNEKODEENFVRELKEMEFLDE- 1391
 DB 759 -----EESCKRKL-----ITLIKVELLNQMKVADOKITALTGSKINIMWDR 802
 QY 1391 -----IKMEASEVEKLFEGKALPIR 1409
 DB 803 ITVAFCKKEFPIEKVFNRSLRK 826

RESULT 13
 OS1465
 ID OS1465 PRELIMINARY; PRT: 2166 AA.
 AC OS1465;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DE 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 GN BB0512.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35210 / B31;
 RX MEDLINE: 98065943
 RA FRASER C.M., CASJENS S., HUANG W.M., STUTTON G.G., CLAYTON R.A.,
 RA LATHIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKRY E.K., GWINN M.,
 RA DOUGHERTY B., TOME J.-F., FLEISCHMANN R.D., RICHARDSON D.,
 RA PETERSON J., KERLAUGE A.R., OJACKENBUSH J., SALZBERG S., HANSON M.,
 RA VAN VUIG R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
 RA UTTERBACK T., WATTHEY L., MCDONALD L., ARITACH P., BOWMAN K.,
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
 RA SMITH H.O., VENTER J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 burgdorferi."
 RL Nature 390:580-586(1997).
 DR EMBL: AE001153; AAC66876.1;
 DR TIGR: BB0512;
 KW Hypothetical protein.
 SQ SEQUENCE 2166 AA; 254243 MW; F2899A0A CRC32;

Query Match 1.9%; Score 144; DB 2; Length 2166;
 Best Local Similarity 16.8%; Pred. No. 5.1;
 Matches 268; Conservative 275; Mismatches 556; Indels 500; Gaps 71.

QY 44 RSNNGYFNSFPLFTSYKTSSTSS-DSHCKDKQSOISIDTSFEIQRFLDKGM---TLI 98
 DB 482 RFNNYNNLNDNINAKTALFESLSSSSKFNQMSKYSKFTD---KLTAGMEFSLM 537
 QY 99 LEKOMROFIOLATVYCTFVIVPRVADVALKTCACLTACRELELAKCIANPSCANAVAC 158
 DB 538 YGEKFETLSQ-----EATNNYQEPDNLKRLNE-----567
 QY 159 LQTCNNRPDETEQIKCGDLFENSVDQFNEGAVSRKRCVPRKSDVGEFVDPDRNAVQON 218
 DB 567 IESFYNNFEETQETLK-----VD-FNISLIN-----IKDELGR-----599
 QY 219 FNNKDESGKMYITSGNPTFADFQCLHEFHTEGDKLVGNISWRIKTLDSGEFT 278
 DB 599 -NIVEFRDRY-----DEVNIFVYQLEBSKLOYSKMOGEMDSNLKNI-----641

QY 279 FVQDDPLDGLALYNHNEFLH-----YODDWYILSSQIENKPPDYIFYYYGRANDAMDGYCG 334
 DB 641 -----SQINKTNEEFLSLQIQKDKGIELSSVFNLDSDI-----QKKAIDMHS 686
 QY 335 ---SVIYTRSPTL-----PEIIPNLOKAASVGRDPNNEITTDNSCGPEPLVERLE-- 385
 DB 687 WKDELINLANKSLDLIKVSSBELSSATLKIESLEKVN-----DREYV 730
 QY 385 ---KTAEEGKILLKEAVE-----IEEVEKEVEKVRDPMTLFQALLBGF 427
 DB 731 LKTKGDI-ESLVIEKYEKELKDMQSYSDDEALLGKIFINROEIIKDSVFMLEDLNRKF 789
 QY 428 KELQDEENY-----RELSEKEKELINELQEMATE-VERLGRALPIRLMA 475
 DB 790 -----DQKNFVISKIECDYKLFKEIESEDILNNEKSDLNFIES-----KLDIV 836
 QY 476 LAPH-----NFLANHETIKYVYKSLPGHKRFSMGWEDYFGSIYV--AK 518
 DB 837 SNKSDNQKQIDFLDRISKDILNRKDSINNEVDSKL-----DWOSKLNELTYAKEN 889
 QY 519 ICGSRIP-----RYFRKSPRICGLDSRGQLFSHGK-----552
 DB 890 LSSGKVDLDLDEYTTIKIKELKFSIES-----LESYIEKIDEPNOCAYISDELLQ 943
 QY 552 ---HNLSPAHISINQVPRG-----NSGCKFPKDYALMWMKMQQPAK-TAIVAFILSV 601
 DB 944 DINHNRKRETRLEERLISKFAVLNNESEFVEKVDLSLDKRDIDASFQANIDITLSD 1003
 QY 602 ASKADAVDALKTCTCLLECHRELAKCISN-PACANAVAC-----641
 DB 1004 NKRFNDIN-----KEINGKNEYISNRYGSENISSKLEWEIHEIENLSRRLTOR 1054
 QY 641 -----LQTCNNRPDETEQI-----KCGDLE-----NSVYDFNCAVSR-K 677
 DB 1055 IDLSKGDNDENLOKLKESFDVSKYQVEKELKAYDLTDGSEAKINLIVEIEQYSRLE 1114
 QY 678 KCVPRKSDVGPFPDPDVSVALQKDMKDFSGKWTITGTLPTFADQCLHEFHTEGDKLVGNISWRIKTLDSGEFT 1162
 DB 1115 EADYTRIT-----DNDIM-----QAKERGE--ITNELKNNIESSEPLNDLYKEREFL 1162
 QY 738 V-GNLSWRIRTPDGGFPTRS--AVOKFVODPKPYGLLYNHNDYLYQDDWYILSSQIENKPPDYIFYYYGRANDAMDGYCG 1206
 DB 1163 IESNFEERYST-----FLIESEGAISS-IRDELYKTLTSDEN-----LOIKISE 1206
 QY 795 SPEDYIFYYYKGRN-----DAMDGYGSVLYTRSAVLPESITPELOTAQAKV 841
 DB 1207 MDQNFELIEORSKDIIEFEKELODKIKDCY-GFINSQFGEIKAGVEINIKNHFVCIRKV 1265
 QY 842 GRDNFPIKTDNMGPEPLVERLE--KVEEGERTIIEV-ELIEEVEKVRD---KE 894
 DB 1266 ---NTLI-DDDIYKENEELHRIKIDSKISTEFTSIEKLNLDKVGSGCIDKIANDFNKY 1320
 QY 895 VTLFSKLFBEFKLODEENFLRELSEEMVDYDLGKMEATEVEKLFGRALPIRLMAVA 954
 DB 1321 IELBEKONBSQNLNENKIDNKIKAIKDNLALSOYDLLEKAYADMDEFSEER--NSYIATL 1378
 QY 955 THOFTSPCHDIRFESSDDGIGRLGITRRKING-----IFLKIILPPIQ 998
 DB 1379 SEEFKSSNKEML--FELBSQKLNKLNESDLANNVEKDYIRIKESYHNVSHTKLEEDF 1436
 QY 999 SADIARTGGRSSRPISARSGFSKGFIDVLPFSKNEIKELTAPLKLKLVGLVACALIV 1058
 DB 1437 FKDLKRGELKYSLENFIASYNKIQMLEYLSKN-----1473
 QY 1059 PSADAVDALKTACCLKCRILELAKCIANPACANAVACLOTNNRPDETEQIKCGDLE 1118
 DB 1473 -----LENKTELIOFRLDIOKKKDKKENFYLDTFEFSKKKDMOSEIA--LME 1521
 QY 1119 NSV---VDEFNCAVSRKRCVPR-----KSDLGEFPAPDSVYVONENISDF-----NG 1164
 DB 1522 TITGKRVDEFDVFNWNOISIDSWFLNIKDDVAKDMQEKSTIEKININAEIGISFEED 1581
 QY 1165 KWIITSGNPTFADFQCLHEFHTEGDN-----KLVGNISWRIKTLDSGEFT 1211

Db 1562 IFNFKIGLESKDGEIKAEIETFSNLONEAKKIEQSVHLEFKNIG- SLNKLVLDELKRFV 1640
 QY 1212 RSAAQKFEVDNQ--PGVLYHNDNEYLHYODD-----WYLLSKIEKNPE 1254
 Db 1641 DFKLEKIDENKKTEDLLIOAEVKFLQCKDEDKIFELNOKIEHFTLLSLDLVVR 1700
 QY 1255 DYIFVYVGRNDAMDYGAGAVVYTRSSV-----LPSNIIPELEKAKSIGRD 1301
 Db 1701 EMVDVSISSDK---ESEFQQLIELINKNISFSEKISLYRNNIETISINEVNSFSKISKD 1756
 QY 1302 FSTIRIDNTGCPPALVERIEKTEVEGERTIYKEVEIEBEVEKVEKGRTEMILFQ- 1361
 Db 1757 LG-----LLEDELKSKLSKHS---TSEIETIKSGLOEQIDK-----FEV 1791
 QY 1361 RLAEGFNELKODEENFYRELKSEEM-----EFLDEIK 1392
 Db 1792 EFKKNHKLKLEVDNNLLELSKILNCDVOFNKRISEIK 1830

RESULT 14

ID 096275 PRELIMINARY; PRT: 1558 AA.
 AC 096275;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE RESA-H3 ANTIGEN.
 GN PFB0915W.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
 RA KOONIN E.V., SHALLOM S., MASON T., YU K., FUJII C., PEDERSON J.,
 RA SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERREA M.,
 RA SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
 RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
 falciparum";
 RL Science 282:1126-1132(1998).
 DR EMBL: AE001424; AAC71972.1;
 SQ SEQUENCE 1558 AA; 175658 MW; DA9C2605 CRC32;

Query Match 1.9%; Score 143.5; DB 5; Length 1558;
 Best Local Similarity 17.9%; Pred. No. 3.5;
 Matches 206; Conservative 153; Mismatches 377; Indels 413; Gaps 50;

QY 347 SIIPNLOK-AAKSYGRFN-----NFTTNSCGPEPLVE-----RLEKTAEBG 390
 Db 626 TILNEVETTAESVTFPSNLIEIOENTITNDT---EERLEEHENVLSALENTQSEE 682
 QY 391 EKLKFAVEIEEVEKEV-----EKVRDTEMILFORLGEFKELOD-----BENVREL 441
 Db 683 EK---KEVIDIEIEVKEEVATLILEYQAEESASTITLIFENLENAVESNENAENL 739
 QY 442 SKKEKILNELOMEATEVERLFGRALPIRLMALAFHSNFLANHETIKYVSKLPGRK 501
 Db 740 EKLNEIYFNVLDEKVEITVEISGESLENNMDKAF-----FSEIFDVAKIOENLLTG-- 793
 QY 502 RFSMGWEDYGSIVAKICSSRRIPRYFKSPRCCGLDSRGQLTSHG---KHNLSPA 557
 Db 793 -----MRSIETSIIVISSE--KYVDLENVYSSILDN--IENKKEGLINKLENISS 840
 QY 558 HSIQNAVPGK-NSGCKPFPDVALVMWKGOFAKTAIYAFILISVAKADAVALATCTC 616
 Db 841 EGVQETVTEHVEQVYVDVDPAMK-----DQF-----LGILNEAG-- 878
 QY 617 LKECRLELAKCISNPACAAVACLOTNNRPDETECQ---IKCGDLFENSVDDEFNEGA 673
 Db 878 -LKEMFNLEDEVFS---ESDVIITVEIKDEPQKREVEKTEVILIEEMENIYDVLEE-- 932

QY 674 VSRKKCVPRKSDVD-EPVPDPSVLVQKFDKDFSGKWFITRGINPTFADFCOLHEFH 731
 Db 932 -----EKELTDKMDIAVESEIETISS-DSKET-----ESI 961
 QY 732 TEENKLVGLSKSRITPDGCGFTTSRQVQVQPKYPOGILYHNDNEYLHYODDYILSSK 791
 Db 962 KDKEDVSLVSEVQDND---MDESVEKVL-----LKNMEELMKDAVEINDITSK 1010
 QY 792 -VENSPEDIYFYVYGRNDAMDYGGSVLYTRSAVLPESI--IPELOAAKVRDFTNF 848
 Db 1011 LIEETQF-----LNEVEADLIKMEKLEKLA--LSEDSKEI 1046
 QY 849 IKTDNTGPEPPYVERLEKVEEGE--RTIIEVEEI-----EEVEKYRD-----K 893
 Db 1047 IDAKD-----DTLEKVEIEEHDTITTDVEVELKDVEDKIEKSDLDKLEBDLK 1097
 QY 894 EV---TLFSKLFEGFKELQDNEFNLELSKEMQVLDGLKMEATEVEKLFGRAL---- 946
 Db 1098 EVKEIKELSELEEDYKELKTIEDTLEKKEIEKDHFEKEFESEAEIKDLEADILEVS 1157
 QY 946 -----YFRKLMVATHCFTSPCHDRIRFSSDDGIGRLGITRKINGTFLKIL 994
 Db 1158 SLFEVEEKKLEVEVHELKEVEHITISGDH-----IKG----- 1190
 QY 995 PRIGADLRTTGGRSSRPIISAFSGFSKGFIDYVLPSPKNEKELTAPLLKLVGLACA 1054
 Db 1190 -LEEDDLEEVNDLKGSTLDMKGMELGDMD-----KESLEEDVTAL----- 1231
 QY 1055 FLIVSADAVDLAKTICACILKSGRIELAKCIANPACAAVACLOTN--NRPDETECQIK 1112
 Db 1231 -----GERVESLKDVLSSALGDEBQMKTRKKAQPKLEVEYLK 1269
 QY 1113 CGDLFENSVDENFNECAVRKCKVPRKSDLGEPAPDPSVLVQNFNLSDENGKYYITISGL 1172
 Db 1270 -----EEVKEEPPKKIITKR-----VAFDIK-----K 1292
 QY 1173 NPTFADFCOLHEFHTEGDNKLVGNISWRITLDSGFTTSRQVQVQPKYPOGILYHND 1232
 Db 1293 EPRKDIIVEYEMKDEDIDED-----IEBDEVED----- 1320
 QY 1233 NXYLHYODDWTLLSKINKKPEDYIFYVYGRNDAMDYGAGAVVYTRSSVLPNSITPELE 1292
 Db 1320 -----IEEDKVEDIDEDID-----E 1334
 QY 1293 KAKSIGRD---FSTFRTDNTGCPPALVERIEKTEVEGERTIYKEVEE-----IE 1341
 Db 1335 DLEDIDGEDKDVIDLIYQKEREKIEKVEKKKLEKAYEBSGKLKHVDVMYVQKID 1394
 QY 1342 EEEVEKEV--GRTKMLFORLAEGFNLKODEE---NFRVRLSKEMEFLDEIKM 1393
 Db 1395 KEVDREVSALSKKNDV-----NVLKQNDFFSKVKNFKKYYKVFAPRISAVAA 1445
 QY 1394 EASEVEKLF 1402
 Db 1446 FASIVVGFF 1454

RESULT 15

ID 092150 PRELIMINARY; PRT: 1819 AA.
 AC 092150;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE CAG ISLAND PROTEIN.
 OS Helicobacter pylori 399.
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 RN [1]
 RP SEQUENCE FROM N.4.
 RA AIM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
 RA SMITH D.R., NOONAN B., GUTILD B.C., DEJONGE B.L., CARMEL G.,

RA TUMANO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
 RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
 RA TRUST T.J.:
 RT "Genomic Sequence Comparison of Two Unrelated Isolates of the Human
 RT Gastric Pathogen Helicobacter pylori."
 RL Nature 397:176-180(1999).
 DR EMBL: AE001481; AAD06047.1; -
 SQ SEQUENCE 1819 AA; 207505 MW; 8809P345 CMC32;

Query Match 1.9%; Score 142.5; DB 2; Length 1819;
 Best Local Similarity 19.0%; Pred. No. 4.9;
 Matches 267; Conservative 199; Mismatches 524; Incls 417; Gaps 66;

QY 178 LFENSVVDQFNECAVRKKCVPRKSDVGEFPVDRNAVQNFNMKDFSGKWITSGLNPT 237
 DB 244 LFSRSIFHYF-----VPLEDKSSRFS--KDRNLVYND----- 274
 QY 238 FDAFDQQLHEFHENDKIVGNLTWRKTLTDGGFFTSAVQTFVODPDLGALYNHNEFL 297
 DB 274 -----EQIIRQ--EYNRLKERNEKGNMLDKLF-----FNDDPN--RTLYN----- 312
 QY 298 HYODDYILSSQJENKPDYIFVYRGRNDAMDYGGSVYTRSPITLPSIIPNLQKAAR 357
 DB 312 -----YLNIAEIEDK--NPLRAFYECISN-----GGN--YECECLKLID--KKLQDQMK 354
 QY 358 SVGRDNNFTTNSCGPEPLVERLEKTAEBEGR--LLIKEAVEL-----EEHV 405
 DB 355 KTLAANDCI--KNAKTEERIKCLDLIKDENLKRSLNQQOVVYALDCLNAKTDEER 411
 QY 406 EKEVEKVRDEM--TLFORLEGEKLEODEENFRELSEKEE--EILNELQMAETEVEKLF 463
 DB 412 KEELKLINDEIREFKRELQELQYKDCIKNAKTEAKNECKLGKSKEA--IERAK 469
 QY 464 GRALPRLKLMALAPHSNPLANHETIKYVGSKLPGHKRFSWGMEDYFSGIYVAKICSR 523
 DB 470 QQALDCLKNAKTDEERKECLKN--IPODLQKELLADM-----SVAYKDCVSR 515
 QY 524 RIRYFRK-----SPRIGGLDSRGQLFSGHKNLSPAHSINQVPR-----GNSG 570
 DB 516 ARNEKEKQCEKELLPEAKKLENOALDCLKNKTDEERKECL--KLPDLOSLDIAKKS 574
 QY 571 CKEPKDVN-----LMWEKMGQFAKTAIVAIFILSVASKADAVDAKTCCTC 616
 DB 575 LKAYKCOASQAKTEAEKKECEKELLTPE-----AKKLEEBAKESVAKYLDVCSQAKT--EA 628
 QY 617 LLEKGC-----RLELAKCISNPACANAVACLOTNNRPDETCQ-----IKGDLFE 662
 DB 629 EKEKECKELLPEAKKLEEK-----KSVRAYLDCVSKARNEAKRECEKELLTPEAKKLE 684
 QY 663 NSVYDEFNECAVRKKCVPRKSDVGEFPVPSVLQKFMKDFSGKWFITRGLNPTFDA 722
 DB 685 NQALD-----CLNAKTDEERKECLDLP-----KDLQKVLAKESVRYLDC 727
 QY 723 FDCQLHEFTEENKLVGNLSWRIRTPDGGFFTSAVQKVFVODPKYPILY----- 773
 DB 728 VSKAKNEAEKKECE-----KLTPE-----ARKLLEAKKSVAKYKDCVLAARNE 772
 QY 773 --NHDMNYILYQDWTILSSKVENSPEDYIFVYKGRNDAMDYGGSVLYTRSAVLPEST 830
 DB 773 KEKQCEKELLTPEAKKLEES--KSVKAYLDCVSKARNEAKRECEKELLTPEAKKLEEA 831
 QY 831 IPELQTAQKVGDRDFTFIKTDTGCEPPLVERLEKVEEGERT-----IIEVEEET 883
 DB 832 KESVKAYKDCVSRARNE--KEQEC--EKLLTPEAKKLEESKSVKAYLDCVSKAKNEA 887
 QY 884 E-EVEKVRKDEVTLSKLEEGKEL-----ORDEENFLRELSEK-----EM 924
 DB 888 ERKECEKELLTPEA--RKLLEBAKESVAKYKDCVSRARNEKQCEKELLTPEAKKLEEN 944
 QY 925 DVLGDKMEATEVE--KLFGRALP--IRKLMAVATHCFTSPCHDRIRFPSSDGIIGRLGI 980
 DB 945 QALDCLNAKTAEAKKRCVADLPKDLQKAVLA-----KESVRYLDCVSKAKNEA 994

QY 981 TRKRINGTFLKLLPPIQASDLTTGGRSSRPISARSGFSKIPDIYLPKSNELKEIT 1040
 DB 995 ERKECE-----KLTPE-----EARKLLEBAKESVAKYKDCVSRARNE--KEKQCEKELL 1041
 QY 1041 APLKLV-----GVLA-----CAFLVPSA-----DAVDALKTC-- 1071
 DB 1042 TPEARKLLEQEVKSVKAYLDCVSRARNEKQCEKELLTPEARKLLENOALDCLKNKT 1101
 QY 1071 ACLKGRLEL-----AKCIANPACANAVACLOTNNRPDETCQ-----IKGDLFENS- 1121
 DB 1102 EAEKRCVCKLPDLQKVLAKESVAKYLDVCSARNEKKECEKELLTPEARKLLEESK 1161
 QY 1121 -VDFEVECAVRKKCVPRKSDVGEFPAPPSVLYONFNISDENGKWTITSGLNPTFDAF 1179
 DB 1162 KSVAYLDC--VSKARNEAEKKECEKELLTPEARKLEE-----AKESVKAY 1205
 QY 1180 -DCQLHEFTEGNKLVGNISWRIKTLIDSGFFTSAVQKVFVODPNQGVLYNHNEYLHY 1238
 DB 1206 KDC-----VSARANKEKO-----ECEKLLT 1226
 QY 1239 QDDWYILSSKIEKPEDYIFVYRGRNDAMDYGGAVYTRSSVLPNSITPE--LEKA 1294
 DB 1227 PEARKLLEQEVKSVKAYLDCVSRARNE-----KEKQCEKELLTPEARKLLEKO 1275
 QY 1295 AKSIGRDFSTFIKTDTGCEPPLVERLEKTEVEGRITVKEVEEIEEVEKEVEKVGRT 1354
 DB 1276 RQKDKAIKCLK--NADPNRAIMKCLDGLSDEKEL--KYLQERKRAVLDCLKTART 1331
 QY 1355 E-----MTLEFORLAEGF--NELQODEENFRELSEKEE-----MEFLDE----- 1391
 DB 1332 DEERKCONIYSDLIOEIQKKAQNNKON--QLKTERLHQASECLDNDPDTQDAIPIQ 1388
 QY 1391 -----TMEASVEKLF 1402
 DB 1389 CLEGLSDSERALLIGIKROADVEDRIV 1415

Search completed: October 13, 1999, 22:44:45
 Job time: 703 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 1999, 04:01:21 ; Search time 2132.21 Seconds
(without alignments)
2370.082 Million cell updates/sec

Title: US-09-075-375-3
Perfect score: 1589
Sequence: 1 tatttcatgtgttgcagt.....tcttttcgaagcaaaaa 1589

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database: GenEmbl.*

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vl:*
18: em_fun:*
19: em_hcg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pal:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vl:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_da1:*
39: em_da2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1589	100.0	1589	8	NTU34817 Nicotiana t

2	643.8	67.5	1760	8	LSU31462	U31462 Lactuca sat
3	624.2	39.3	1555	8	ATU44133	U44133 Arabidopsis
4	428.4	27.0	13840	8	AC003981	AC003981 Complete
5	426.8	26.9	2884	8	ATU39452	U39452 Arabidopsis
6	162.4	10.2	89493	8	AC006932	AC006932 Genomic s
7	75.8	4.8	7218	5	166494	166494 Sequence 14
8	52.4	3.3	1686	26	E08995	E08995 DNA encodin
9	52	3.3	2554	7	SCYL1202W	Z28201 S.cerevisia
10	50.6	3.2	3811	11	HSR8153	U88153 Homo sapien
11	50.6	3.2	3121	11	HSR8154	U88154 Homo sapien
12	50.4	3.2	4000	7	D83006	D83006 Saccharomyc
13	50.4	3.2	177568	9	HS388M5	Z97055 Human DNA s
14	49	3.1	2222	1	SHTLYCA	X73141 S.hydroysent
15	49	3.1	5001	36	PEAABRA	J03902 Plasmodium
16	48.6	3.1	6152	5	A58933	A58933 Sequence 1
17	48.6	3.1	5361	5	A58933	A58933 Sequence 2
18	48.6	3.1	12240	37	AP001424	AE001424 Plasmodiu
19	48.4	3.0	19486	11	AF055066	AF055066 Homo sapi
20	48.2	3.0	312766	34	PEMALP3	AL034726 Plasmodiu
21	47.4	3.0	12029	37	AE001373	AE001373 Plasmodiu
22	46.6	2.9	11011	2	AE001175	AE001175 Borrelia
23	46.6	2.9	151965	35	AC006454	AC006454 Homo sapi
24	46.4	2.9	334	7	HVC39SAT	Z50100 H.vulgaris G
25	46.4	2.9	86829	36	PEMALP5	AL034556 Plasmodiu
26	46.2	2.9	128650	34	AF127414	AF127414 Homo sapi
27	45.8	2.9	204038	34	AC006371	AC006371 Homo sapi
28	45.4	2.9	800	37	TG053810	U53810 Toxoplasma
29	45.2	2.8	200000	34	AC004670	AC004670 Homo sapi
30	45.2	2.8	52670	34	PEMALP3	AL031746 Plasmodiu
31	45.2	2.8	226683	35	AC007297	AC007297 Homo sapi
32	45	2.8	197913	11	AC005725	AC005725 Homo sapi
33	45	2.8	2069	26	E10126	E10126 DNA encodin
34	45	2.8	3399	26	E10126	E10126 DNA encodin
35	44.6	2.8	36339	10	HSB33B7	Z82176 Human DNA s
36	44.6	2.8	151840	34	HS919B11	AB013102 Homo sapi
37	44.4	2.8	421	9	AB018492	AB018492 Homo sapi
38	44.4	2.8	95565	11	AC000379	AP000379 Homo sapi
39	44.4	2.8	100000	9	AP000067	AP000067 Homo sapi
40	44.4	2.8	95565	11	AC000379	AC000379 Human Chr
41	44.4	2.8	157248	11	AC002312	AC002312 Human Chr
42	44.4	2.8	94882	34	AC002317	AC002317 Homo sapi
43	44	2.8	3740	9	AB018344	AB018344 Homo sapi
44	44	2.8	9833	12	MMV1MENT	Y07738 M.musculus
45	44	2.8	5420	36	AF056936	AF056936 Plasmodiu

ALIGNMENTS

RESULT 1	NTU34817	1589 bp	mRNA	PLN	15-JUN-1998
LOCUS	Nicotiana tabacum				
DEFINITION	Nicotiana tabacum violaxanthin de-epoxidase precursor (1VDE1), mRNA, complete cds.				
ACCESSION	U34817				
VERSION	U34817.1	GI:1463122			
KEYWORDS	common tcbacco.				
SOURCE	Nicotiana tabacum				
ORGANISM	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;				
REFERENCE	1 (bases 1 to 1589)				
AUTHORS	Bugos,R.C., Hieber,A.D. and Yamamoto,H.Y.				
TITLE	Xanthophyll cycle enzymes are members of the lipocalin family, the first identified from plants				
JOURNAL	J Biol Chem. 273 (25), 15321-15324 (1998)				
MEDLINE	98288256				
REFERENCE	2 (bases 1 to 1589)				
AUTHORS	Bugos,R.C. and Yamamoto,H.Y.				
TITLE	Direct Submission				

JOURNAL Submitted (25-AUG-1995) Robert C. Egnos, Plant Molecular Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI 96822, USA

FEATURES
source Location/Qualifiers

1..1589
/organism="Nicotiana tabacum"
/strain="Xanthi"
/db_xref="taxon:4097"
/tissue_type="leaf"
1..1589
/gene="TVDE1"
42..1478
/gene="TVDE1"
/codon_start=1
/product="violaxanthin de-epoxidase precursor"
/protein_id="AAC5003.1"
/db_xref="pid:g1463123"
/db_xref="gi:1463123"

gene

CDS

translation="MALAPHSNPLANHEITIKYVSGKLPGRKFSWGMEDYFGSTIVA
DVALMWEKMGCFAPKTAIVAFILTSVSKADADALKTCCILKECEBLACISPA
CAANVACIOTNNRDETECOIKGDIENSVDDEFNECAVSKKCVKSVGDFPV
PDSVIVOKEMKDPKSKMPTIRSLNTEFADPCQLHEFTENKLVGNLSWRITPD
NDAMDGGGSLYIRSAVLPESIIPELQTAOKVGDFTNFIKTNDTCGPPELYERL
EKVDEGERITIKVEEIEIEVEKVRKREVTLEFSKLEFGFKEIQRDEENFLELSKEE
MDVLDGLKMEATEVEKLEFGRALPRKLR"

transit_peptide 42..443

/gene="TVDE1"
mat_peptide 444..1475
/gene="TVDE1"

/product="violaxanthin de-epoxidase"

BASE COUNT 500 a 262 c 366 g 461 t

ORIGIN

Query Match 100.0% Score 1589; DB 8; Length 1589;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 tattttcatgagtttgagttggtgtaatacgttgaagaatggtcttgcctccatt 60
1 TATTTCATGAGTTTGCAGTTGGTGAATACGTTGAAGAAATGCTTGGCCCTCAT 60
61 caatttctggccaacacatgaacacatcaatattatgttgggtcaagctccggtc 120
61 CAATTTCTGGCCACCAACCAATGAAATATATATGTTGGTCAAAAGCTTCCGGT 120
121 ataaaggtttagctgggtgtggaagatcttggtagtaagcgtgagaataatt 180
121 ATAAAGGTTAGCTGGGTGGGAGATTACTTGTGTAGTATAGTCGTAGCAAAATTT 180
121 gtcccaagcagagatacctaagatcttcgaataatcttcctgaatagtgtgttgg 240
121 GTTCCAGCAGAGATACCTAAGTACTTTCGAAATCTCTCTGAATATGCTGTGTTGG 240
181 attcaagaggtctgcaactatctcacacggaagaacacacatctctccgacataga 300
181 GTTCCAGCAGAGATACCTAAGTACTTTCGAAATCTCTCTGAATATGCTGTGTTGG 240
241 attcaagaggtctgcaactatctcacacggaagaacacacatctctccgacataga 300
241 ATTCAGAGAGTGTGCAACTATCTCTCACAGGAAACACATCTCTCTCCGACATACA 300
301 ttacccaagatgaactaagaggaatctcagaatgcaaatlccaaaagatgtagcttga 360
301 TTACCCAAGATGACTAAGGAAATTCAGATGCAAAATTCAAAAGATGACTTTGA 360
361 tggtttggaagaatggtgccaatttgcacaaagcagcaattgtgctatattcatttgc 420
361 TGGTTTGGAGAAATGGGSCCAATTTGCCAAATACGCAATGTGACTATATTTTGT 420
421 cagttgctcaaaagcgtgagtggttgaatgctcacaagactgtactgtctacttgaag 480
421 CAGTTGCTCAAAAGCTGATGCGGTTGATGCTCAAGACTGTACTGCTTACTGGAAG 480
481 agtgcaggttagagcttgaggaagtgcatttcgaacctgcagtgcagcaatttgcct 540

481 agtgcaggttagagcttgaggaagtgcatttcgaacctgcagtgcagcaatttgcct 540
481 AGTGCAGGTTAGAGCTTGAGGAAGTGCATTTCGAACCTGCATGCGAGCAATAGTGGCT 540
541 gtctccagacttgcaacacatagacctgaacgaacgaatctcagaataatgctggtatt 600
541 GTCTCCAGACTTGCAACATAGACCTGACGAACGAATCTCAGATTAATGTTGTTGATT 600
601 tgttgaagaacagtgctgtagacagatccaatgagtgtgcagctcccgaaagaaatgtg 660
601 TGTTGAAACAGTGCTGTACAGCATGATGATGATGATGATGATGATGATGATGATG 660
661 taactcgtaaaatcgatggtggtgactcttcctgactgactgactgactgactgactg 720
661 TAACCTGTAATCGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 720
721 agtttgaatgaagaattttagcgggaatggttcaatctcgcggttgaatccact 780
721 AGTTTGAATGAAGATTTAGCGGGAATGTTCAATCTGCGGTTGAATCCACT 780
781 ttgagtctttagatgccaatgcatgagttccatcagaagaagaagaagaagaagaaga 840
781 TTGAGTCTTTGATGTCATATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 840
841 attatcttgaggaatacgtacacactgagaggaatttllactcgatcagcggtgcaaa 900
841 ATTATCTTGGAAGATACGTACACTGATGAGGATTTTACTGATGATGATGATGATG 900
901 aattcgtcaagaatccaaagatccgggagatctcacaatcatgataatgatacttc 960
901 AATTCGTCAAGATCCAAAGTATCCGGGATCTCTACATCATATATATATATATATAT 960
961 tctaccaagatgactgtagtatttgcataccaagaatgaataatgataatgataatgata 1020
961 TCTACCAAGATGACTGATATTTTCTCATCAAGTGAAGAAATGTCACAGGATTTACA 1020
1021 tatttgtactaagagcagaagaatgataatgataatgataatgataatgataatgataat 1080
1021 TATTGTGATCTTAAGGCGCAAAATGATGATGATGATGATGATGATGATGATGATG 1080
1081 acaagaagatgaggttcttgcctgaagaacatctacggaatgcaacggcagctcaaa 1140
1081 ACAGAAGATGAGGTTGCTGCTGAAGCATATATACCGGATTTGCAACCGCAGCTTAAA 1140
1141 aatttggcgtgtaattccaacatctcacaataaagaacacacacacacacacacacac 1200
1141 AATTGGCGCTGATTTCAACACATTCATTAACAGACAAATACATGTTGGCTGAACTCT 1200
1201 cccctgttgaagaggttggagaagaagtgagaagaaggaagaagaagatcacaataagaag 1260
1201 CCTTGTGTGAGAGGTTGGAGAAGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1261 ttgagaagaatgagaagaatgagaagaatgagaagaatgagaagaatgagaagaatgaga 1320
1261 TTGAGAGATGAGAAGAAGATGAGAAGTGAAGAGATGAAGAGATGAAGAGATGAAGAG 1320
1321 aactgttgaaggttllaaagagctccaagagatgagaagaactcttgaagaactcga 1380
1321 AACTGTTGAAGGTTTAAAGAGCTCCACAGCATGTAAGAGACTCTTAAAGAGACTGA 1380
1381 gaaaagaagaatgagatgttllgaatgactaaatgagaagaactgagatgagaagaac 1440
1381 GAAAAGAAGAATGAGATGTTTGAATGAGACTTAAATGGAAGCACTGAGGTGGAAGAAC 1440
1441 ttlttggcgtgcttcccaatgaagaatgaagaatgaagaatgaagaatgaagaatgaaga 1500
1441 TTTTGGCGGCTTTACCAATTAAGCAATTAAGGTAAGTATTTTAAATCAATCAATCA 1500
1501 atatactacatgataatgagatgatttgcctcgaatgataatgataatgataatgata 1560
1501 ATATACATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
1561 tcatctgcttcttllagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1589
1561 TCATCTGCTTCTTLLAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 1589

Db 1683 A 1683

RESULT 3
 LOCUS ATU4133 1555 bp mRNA PLN 15-JUN-1998
 DEFINITION Arabidopsis thaliana violaxanthin de-epoxidase precursor (AVDE1)
 ACCESSION U44133
 NID 91465734
 VERSION 044133.1 GI:1465734
 KEYWORDS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.

REFERENCE 1 (bases 1 to 1555)
 AUTHORS Bugos, R.C., Hieber, A.D. and Yamamoto, H.Y.
 TITLE Xanthophyll cycle enzymes are members of the lipocalin family, the first identified from plants
 JOURNAL J. Biol. Chem. 273 (25), 15321-15324 (1998)
 MEDLINE 98288256
 REFERENCE 2 (bases 1 to 1555)
 AUTHORS Bugos, R.C. and Yamamoto, H.Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-JAN-1996) Robert C. Bugos, Plant Molecular Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI 96822, USA

FEATURES

source location/Qualifiers
 1..1555
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 1..1555
 /gene="AVDE1"
 45..1433
 /gene="AVDE1"
 /codon_start=1
 /product="violaxanthin de-epoxidase precursor"
 /protein_id="AAC50032.1"
 /db_xref="pid:91465735"
 /db_xref="gi:1465735"

translation="MAVATHTGSPCHDRIRNFSSDDGICRLGTRKINGTEFLKIL
 PIOSADLRITGGRSSRLSPFRSGFSGITDIVLPSSKMELETPLLKLVGLA
 CAFILPPSDADALKTACALKGCRIELANCPACANVACICTCNRPDETGO
 IKCGLPFNSVVDDEFNECAVRKCVPRKSLDGFPPADPSVYONNPNRPEKCY
 TSGLPNTEADFCLEHFEHEDGNKLVGNISWIKTLDGSEFTSRSAVOKFQPNPG
 VLYNDNEYLYHODDMYLLSKLENKPEYFVYRGANDAMGCGAVYTRSSVLP
 NSIPELEKAKSIGRDSFTIRTDNCGPEPALVERIKTEVEGERITVKEVELEE
 EKEVEKVGRTMLTFORLAEGFNEIKODEENFRELSEMEFLDEIKMESEVEK
 LEKALPIRKVR"

transit_peptide 45..363
 /gene="AVDE1"
 mat_peptide 364..1430
 /gene="AVDE1"
 /product="violaxanthin de-epoxidase"
 BASE COUNT 485 a 285 c 376 g 409 t
 ORIGIN

Query Match Best Local Similarity 39.3%; Score 624.2; DB 8; Length 1555;
 Matches 820; Conservative 0; Mismatches 268; Indels 15; Gaps 2;

QY 424 ttgctcaaaagcttgatcggtgattgctctcaagactgttactgtcttcaaaagt 483
 DB 364 TTGTCCTTCGTCGATGCAATGCACTTAAACCTTGTCATGCTTAAAGAGAT 423
 QY 484 gcaagttgagcttgcaagttgcaagctgcatgtgcaagtaattgttcctgc 543
 DB 424 GCAGGATGAGACTGCAAGTGCATGCCAACCTGCTGTGACCAATGTGGCTGCC 483

QY 544 tccagacttgcaacaatagacctgacgaacgaaatgctcagataaaatggtgattgt 603
 DB 484 TTGACAGCTGCAATACCGTCCAGATGAACCGAGTCCAGATTAAATGGGGATCTGT 543
 QY 604 ttgaacaagctgctgtagcagagttcactgaagtgctcagctcccgaaagaaatgtgtac 663
 DB 544 TTGACAAAGCTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 603
 QY 664 ctgtaaatctgactgt 723
 DB 604 CTGAAATCTGATCTCGAGAAATTTCCGCCCAACCTTCTGTTCTGTTGACAACT 663
 QY 724 ttgacatgaagatttgaaggggaatggttcaactctgcggtttgaatccactttg 783
 DB 664 TCACATCTCGACTTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 723
 QY 784 atgctttgactgcaacttgatgactgactgactgactgactgactgactgactgactgact 840
 DB 724 ATGCTTCGACTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAG 783
 QY 841 attatcttgagaataacgtacacccgtatgagagatttttactgcacgagtgcaaa 900
 DB 784 ACATCTCTTGAGAAATTAAG 843
 QY 901 aattctgcaagatcccaagatccgggatactctacatcagatgataatgattatcttc 960
 DB 844 AATCGTGCAAGATCTTAACCAACCTGCTGTTCTTAACATCATGACAAACAGTACCTTC 903
 QY 961 tctacaagaatgactgattatattgcaatccaaagtagaataatgctcagagattaca 1020
 DB 904 ACATCAAGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 963
 QY 1021 tatttgactactaagaagcagaatgactgactgactgactgactgactgactgactgact 1080
 DB 964 TATTGTATCTACCGTGGCGCAACGATGCTTGAGAGATGATGATGATGATGATGATGAT 1023
 QY 1081 acgcaagaagctgactgactgactgactgactgactgactgactgactgactgactgact 1140
 DB 1024 AACGGAAGATGCTGATTAATACCAATACCAATACCAATACCAATACCAATACCAATAC 1083
 QY 1141 aagttgggagttatccaacatccaatcaaaagcaatcaatgctgactgactgactgact 1200
 DB 1084 GATAGGAGAGATCTTACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1143
 QY 1201 cccctgttgagaagtttgagaagaagaatggaagaagaagaagaagaagaagaagaaga 1260
 DB 1144 CCGTGTGAGAGAGATGAG 1203
 QY 1261 ttgaagagatagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1308
 DB 1204 TTGAAGAGATGAG 1263
 QY 1309 ccttattcaagaaatttgagaagtttgaagaagttccaagaagttgaagaagttct 1368
 DB 1264 CTTGTTCCAGAGATTTGCTGAGAGATTTAATCAATCAATCAATCAATCAATCAATCAAT 1323
 QY 1369 taagaagagctgagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1428
 DB 1324 TGAGAGAGATTAAG 1383
 QY 1429 agttagaagaacatttgagcggtgcttccatgaagaagaagaagaagaagaagaagaaga 1488
 DB 1384 AGGTGAAATATTGTTGGGAAAGCTTTCCATCAAGAGAGAGAGAGAGAGAGAGAGAG 1443
 QY 1489 aactatcaatataatacaat 1511
 DB 1444 CACATTTGTTGACAAACTATAT 1466

RESULT 4
 AC003981 133840 bp DNA PLN 18-SEP-1998

Q079J937 ~ similar to DNA dependent Atase and helicase

Query Match	27.08;	Score 428.4;	DB 8;	Length 133840;
Best Local Similarity	67.68;	Pred. No. 1.6e-79;		
Matches 703; Conservative	0;	Mismatches 331;		

OY	578	atctcagataaataatggttgattgttttgaacaagtcgcagacagatcaatgagtc	637
Db	6304	ATATCAGATTAAATATGTGGGAAATCGTTTGAGAACACTGTGTGATGATCAACGAGTC	6366
OY	638	tgcagctcccgaaagaatctgacctgctaattgatgtgtgtaacttccttacc	697
Db	6364	TGCTGTGTGCAAAAAAGATGTGTCTTAAATAATGTGATCGGAAATTTCTCTCCC	6423
OY	698	tgaatccagatgtctctgtgccagaagtggacatgaagaatcttcgcggagaatcgtctat	757
Db	6424	AGACCTTCTGTCTTGTGAGAGAACTCAACAATCGGACTTTAACGGAGAGTGACT	6483
OY	758	tactcgcggtttgaatcccaactttgatgtctttgatgtgcaaatctgaatgagtctcaac	817
Db	6484	TCAAGTGGCTTGATCAACACTTGTATGCTCCGACGCGAGTGCATGAGTGTCCACG	6543

QY	818	agaa-----gaaacaacaacttgtyggggaattlacttggagataacgtaaccctgatggag	874
Db	6544	AGAAAGTGAAACAACAACCTGTGTGGAAACATCTCTTGGAGATAAAGACCTTAGAACGTGG	6603
QY	875	atttttaccgcgatcagcgggtgcgaaaatttgctcaagatccaaagatcccgggatact	934
Db	6604	ATTCTTTACTAGGTCAGCCGCTACAAAATTTGTGTCAAGATCTTAACCAACCTGGTGTCT	6663
QY	935	ctaaacatcatgaataatgagatctctcttaccaaatgact-----	976
Db	6664	CTAACAATCATGAACAACGATTAACCTTAACAGATGGTGTAAATCACTTTGTATC	6723
QY	976	-----	976
Db	6724	CACCCACAGGAGAGCTGTGATAGTCTGTATCTTCTTTGGCTTAACAAGCTTTCAT	6783
QY	976	-----gtaataattgtcatccaagtagaaatgtccaaagattacat	1023
Db	6784	ATGATTTTGCAGGTATATCTGTCTATCAACAAGTATAGAAATTAACCTTAACATATAT	6843
QY	1024	tttgtactataagggcgaatgaatgcatgagatgatatgtgtctgtacttaca	1083
Db	6844	TTGTATACTACCGTGGGGGAACCATCTTGGATGATATGTGTGTCACTTGTATACA	6903
QY	1084	caagaagtgcgaatttgcgtgaaagatataccggagttgcaaacgcgcgtccaaaag	1143
Db	6904	CGAGAAAGTTCTGTATATCCCAATATGCAATTAATCCAGAACTCGAAAAAGCAGCAAAAGCA	6963
QY	1144	ttggcgcgtgatccaacacatlataaacaacagacatatactgtggccctgaacccccc	1203
Db	6964	TAGGCGAGAGACTTCAGACATTCATTAGAACGGATTAACACTGTGTCTCGAACCTGGCC	7023
QY	1204	ttgttggaaggtttggagaagaagtgtgaagaaggagaagggatcatlaaagaagtgtg	1263
Db	7024	TGCTGGAAGCATTTGACAAACACAGCTGGGAAGAGTGGAAGATATGTGTAAAGAGCTTG	7083
QY	1264	aggagatagaagaagaagtgtgagaaggtgtgagataa-----agaatcaact	1311
Db	7084	AAGAGATAGAAAGAAAGGTAGAGAACGAAGTGGCAAGTGTGGTAGACTGAGATGACT	7143
QY	1312	tattcagtaaacgtgttgaaggttttaaagagctccaacgagatgagaagaactcttaa	1371
Db	7144	TGTTCCAGAAATTTGGCTGAGGATTTATTAAGCAATGAGCAAGACGAGGAATTTCTGGA	7203
QY	1372	gagagcgtgagcaagaagaagaatgsgtgttttgatgtgaacttaaatgtgaagaacttgag	1431
Db	7204	GAGAGTTAATTAAGACAGAGATGAGGTTTGTGATGATCAAAATGCAAGCAAGTAGAG	7263
QY	1432	tagaaaaacttttggcgtgtgttaccaaatgaagaatlaagatgaatattttaaac	1491
Db	7264	TTGAAAATTTGTTGGGAAGCTTTGCCATTCAGGAAGGTCAAGTTCAGTGAACAGAACCCAC	7323
QY	1492	tatcaacatataactaat	1511
Db	7324	CATTGTTGTRCAAACTATAT	7343

[illegible]

353 ccgcaacccgagcgcacaaaggctgggcgtgattcaacacattcataaaacagac 1178

[illegible]

```
/db_xref="PID:g2980817"
```

1


```
repeat_region /chromosome="22"
/clone="388M5"
/clone_id="RPI13"
318. .452
/notes="Alu repeat: matches 1. .136 of consensus;
incomplete repeat"
repeat_region 2291. .2413
/notes="MIR2 repeat: matches 145. .23 of consensus"
repeat_region 3079. .3424
/notes="MER1B repeat: matches 337. .1 of consensus"
repeat_region 3505. .3681
/notes="MIR repeat: matches 248. .71 of consensus"
repeat_region 5509. .5657
/notes="MIR repeat: matches 101. .262 of consensus"
repeat_region 5805. .6105
/notes="MIR repeat: matches 101. .262 of consensus"
repeat_region 6733. .6842
/notes="Alu repeat: matches 301. .2 of consensus"
repeat_region 7891. .8064
/notes="MIR repeat: matches 160. .47 of consensus"
repeat_region 9675. .9769
/notes="MIR repeat: matches 218. .31 of consensus"
repeat_region 9774. .10005
/notes="MIR repeat: matches 258. .7 of consensus"
repeat_region 10956. .11256
/notes="Alu repeat: matches 2. .303 of consensus"
repeat_region 11869. .11951
/notes="LI repeat: matches 2204. .2120 of consensus"
repeat_region 12630. .12789
/notes="Alu repeat: matches 1. .160 of consensus;
incomplete repeat"
gene /gene="GJ388M5.1"
complement(12954. .13560)
CDS /partial
/gene="GJ388M5.1"
/notes="60S Ribosomal protein L1 like pseudogene: match:
proteins P49626 Q28346 P09180 P36578 P50878 P49691 P08429
P02385 Q02056 P49165 P14117"
/pseudo
/evidence="not_experimental"
/db_xref="PID:e1319651"
repeat_region 13613. .13912
/notes="Alu repeat: matches 2. .302 of consensus"
repeat_region 14345. .14614
/notes="Alu repeat: matches 302. .38 of consensus;
incomplete repeat"
repeat_region 14625. .14710
/notes="L1M2 repeat: matches 553. .639 of consensus"
repeat_region 14625. .14692
/notes="L1M8 repeat: matches 553. .621 of consensus"
repeat_region 14739. .14772
/notes="Alu/g repeat: matches 34. .1 of consensus;
incomplete repeat"
repeat_region 14814. .14994
/notes="MST repeat: matches 217. .394 of consensus"
repeat_region 15333. .15566
/notes="3 copies of 78 mer 94 & conserved"
repeat_region 16733. .17027
/notes="Alu repeat: matches 295. .1 of consensus"
repeat_region 17031. .17202
/notes="L1M2 repeat: matches 915. .737 of consensus"
repeat_region 17459. .17750
/notes="Alu repeat: matches 302. .1 of consensus"
repeat_region 17751. .18336
/notes="L1P16 repeat: matches 902. .292 of consensus"
repeat_region 19526. .19716
/notes="MER42C repeat: matches 1185. .1386 of consensus"
repeat_region 19725. .20021
/notes="Alu repeat: matches 300. .1 of consensus"
repeat_region 20085. .20184
/notes="MER42C repeat: matches 1398. .1494 of consensus"

repeat_region 20220. .20502
/notes="Alu repeat: matches 294. .1 of consensus"
repeat_region 20345. .20656
/notes="Alu repeat: matches 302. .193 of consensus;
incomplete repeat"
repeat_region 22630. .23196
/notes="L1M7 repeat: matches 333. .918 of consensus"
repeat_region 23480. .23536
/notes="Alu repeat: matches 255. .199 of consensus;
incomplete repeat"
repeat_region 23546. .23845
/notes="Alu repeat: matches 298. .1 of consensus"
repeat_region 23851. .23989
/notes="Alu repeat: matches 140. .1 of consensus;
incomplete repeat"
repeat_region 24838. .25133
/notes="Alu repeat: matches 1. .299 of consensus"
repeat_region 25854. .25966
/notes="MIR2 repeat: matches 22. .141 of consensus"
repeat_region 26047. .26350
/notes="Alu repeat: matches 1. .303 of consensus"
repeat_region 27255. .27552
/notes="Alu repeat: matches 1. .299 of consensus"
repeat_region 29971. .30060
/notes="match: ESTs T95489 T95490"
repeat_region 29138. .29432
/notes="Alu repeat: matches 297. .1 of consensus"
repeat_region 29644. .29759
/notes="MIR1F repeat: matches 442. .568 of consensus"
repeat_region 29675. .29759
/notes="MIR1F repeat: matches 458. .541 of consensus"
repeat_region 29971. .30060
/notes="MIR2 repeat: matches 146. .59 of consensus"
repeat_region 33150. .33448
/notes="Alu repeat: matches 2. .299 of consensus"
repeat_region 33563. .33736
/notes="L1M1 repeat: matches 12. .194 of consensus"
repeat_region 33894. .34195
/notes="Alu repeat: matches 1. .302 of consensus"
repeat_region 34685. .34818
/notes="MIR2 repeat: matches 7. .146 of consensus"
repeat_region 34910. .35008
/notes="L1M2 repeat: matches 397. .496 of consensus"
repeat_region 35011. .35318
/notes="Alu repeat: matches 302. .1 of consensus"
repeat_region 35320. .35581
/notes="L1M2 repeat: matches 817. .1075 of consensus"
repeat_region 35619. .35814
/notes="MIR repeat: matches 32. .236 of consensus"
repeat_region 37465. .37776
/notes="Alu repeat: matches 1. .302 of consensus"
repeat_region 38058. .38192
/notes="FLAM C repeat: matches 1. .133 of consensus"
repeat_region 38233. .38400
/notes="4 copies of 42 mer 88 & conserved"
misc_feature 38254. .38260
/notes="single clone sequence"
repeat_region 38408. .38592
/notes="Alu repeat: matches 125. .302 of consensus;
incomplete repeat"
repeat_region 38685. .39047
/notes="LI repeat: matches 2437. .2810 of consensus"
repeat_region 39281. .39750
/notes="LI repeat: matches 3062. .3546 of consensus"
repeat_region 39770. .40006
/notes="Alu repeat: matches 1. .237 of consensus;
incomplete repeat"
repeat_region 40020. .40128
/notes="LI repeat: matches 3573. .3686 of consensus"
repeat_region 40129. .40419
/notes="Alu repeat: matches 1. .301 of consensus"
repeat_region 40435. .42181
/notes="LI repeat: matches 3663. .5390 of consensus"
```


Human pp32 cDNA. N
Continuation (17 0

(without alignments)
2169.943 Million cell updates/sec

Title: 02-07-2016
Perfect score: 1589
+-----+
tcttttcagaagcaaaaa 1589

Scoring table: IDENTITY_NUC

Database : N_Geneseq_36:*

SUMMARIES

Query Match	Score 1587.4	DB 1	Length 1589
Best Local Similarity 99.9%	Pred. No. 0		
Matches 1588	Cons. Ivalative 0	Mismatches 1	Indels 0
			Gaps 0
1 tattttcattggtttgagctgtgtgtaatacgggttgaagatgtgctcttggccctcatt	60		
1 TATTTTCATGAGTTTCCAGTGTGTGTATACGCTTGAAGATGCGTCCCTCATTT	60		
61 caaatcttcctgccaacatgaacatcaatataatattgttgggtcaaaagttccggtc	120		
61 CAAATTTTCTGGCCAAACATGAACATCAATATATATGTGGTCAAAAGCTTCCCGTCC	120		
121 ataaaggtttatctgggttgggaagattactttgtgtagtagtcgttagcaaaaatt	180		
121 ATAAAGGTTTATCTGGGGTTGGGAAGATTACTTGTGTAGTATGCTAGCAAAAATTT	180		
121 gtccagcagacagatatacctagatacttccgaaaatcctccagaataatgctgtgttg	240		
121 GTTCCAGCAGACGATACCTAGATACTTTGAAAATCTCCTCAATATCTGTGTGTGG	240		
181 gtccagcagacagatatacctagatacttccgaaaatcctccagaataatgctgtgttg	240		
181 GTTCCAGCAGACGATACCTAGATACTTTGAAAATCTCCTCAATATCTGTGTGTGG	240		
241 attcaagagctctgaactattctccaagggaaacaacaatctctccgcacatagca	300		
241 ATTCAAGAGCTCTGAACCTATTCTCACACGGGAAACCAATCTCTCCCGCACATAGCA	300		

QY 1381 gcaagaagaatggatggtctttggatggaacttaaatggaagcaattggtggaanaac 1440

Db 1381 GCMAAGAAAGAAATGGAGTGTGGATGGACTTAAATGGAAGCAACTGAGTGAAGAAAAC 1440

QY 1441 tttttggcgcttcctaccataagaanaatgaagcgaatcttttaaacatataacat 1500

Db 1441 TTTTGGCGCTGCTTTCACAAATAGGAANAATTAAGTAAGTATTTTAAACATACACAT 1500

QY 1501 atatacaaatgatatagttgatatctttgattccttcctggaataatgctataacat 1560

Db 1501 ATATACTAACATGATATGTTGATTTATTTATCTTTTCCCGAATACATTTGCTTATACATCA 1560

QY 1561 ttttatgctctctcttcagaagcaaaaa 1589

Db 1561 TGTATGCTCTCTTTTCACAGCAAAAAA 1589

```

RESULT      2
ID          T66241
AC          T66241; standard; cDNA; 1981 BP.
DT          28-JUL-1997 (first entry)
DE          Romaine lettuce violaxanthin de-epoxidase cDNA.
KW          Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW          photoprotection; transgenic plant; zeaxanthin; anthexanthin;
KW          xanthophyll; lettuce; ss.
OS          Lactuca sativa L. cv. Romaine.
FH          key
FT          misc_difference 26..29
              Location/Qualifiers
FT          /tag= a
FT          /note= "bases 26-29 are illegible in Fig 1"
FT          misc_difference 66..72
              /tag= b
FT          /note= "bases 66-72 are illegible in Fig 1"
FT          misc_difference 105..110
              /tag= c
FT          /note= "bases 105-110 are illegible in Fig 1"
FT          misc_difference 147..149
              /tag= d
FT          /note= "bases 147-149 are illegible in Fig 1"
FT          misc_difference 186..189
              /tag= e
FT          /note= "bases 186-189 are illegible in Fig 1"
FT          misc_difference 226..227
              /tag= f
FT          /note= "bases 226-227 are illegible in Fig 1"
FT          cds
              /tag= g
FT          /note= "bases 235-1656 are illegible in Fig 1"
FT          translat_peptide 235..609
              /tag= h
FT          /note= "bases 610-1653 are illegible in Fig 1"
FT          mat_peptide
              /tag= i
FT          W09717447-A2.
PE          15-MAY-1997.
PF          07-NOV-1996; U18291.
PR          07-NOV-1995; US-006315.
PA          06-AUG-1996; US-023502.
PI          (CALJ ) CALGENE INC.
DI          Bugos RC, Rockholm DC, Yamamoto HY;
DR          WP1; 97-281036/25.
P-PEDB:     W09874.
PT          DNA encoding plant violaxanthin de-epoxidase - used to modify the
PS          sensitivity of a plant to light
PT          Claim 3, Fig 1; 41pp; English.
CC          A cDNA clone (T66241) codes for the 55 kDa violaxanthin de-epoxidase
CC          (VDE) (W09874) of romaine lettuce. VDE was purified from romaine
CC          lettuce chloroplasts and 2 tryptic peptides were used to develop
CC          primers (see also T66244-45), which amplified a partial VDE
CC          sequence. The amplified bp DNA sequence was then used to screen a lettuce
CC          cDNA library, and the 1881 bp DNA sequence was identified. VDE
CC          nucleic acids (see also T66244-43), in sense or antisense
CC          orientation, can be used in genetic constructs, pref. also contg. a
CC          plastid translocation sequence, to modify VDE levels in plants.

```

Oy	317	taaggnaaatlcagatgcaaatctccaaagaatgcttgattggttggaagaatg	376
Db	483	TCAGCAAAATACAAAATTTGATCTCAAAAGGGGCACTGACTTTGATCTTGAAAAGCATG	542
Oy	377	gggcgaatttcgcaaacagcaatgtagcatalcatcatttgcagtgcctcaaacg	436
Db	543	GGACAAATTCATACAAATTGGCATCGTATGGTTGGCCATTGTGTATCGTTCCTAAGT	602
Oy	437	tgatcggtttagtgcctcaagaacttgtaacttgcttactgaaagaagtgcaagttagc	496
Db	603	TCATGCGCTGTAGCTCTTAAACCTTGCTGTTACTCAAGAGATCGAGATTGAGCT	662
Oy	497	tgcgaagtcatttcgaaaccgcagctgtagcaatacttgccctgcctccagactgcga	556
Db	663	TGCAAAATTTATAGCAAAACCCATCTTGTCGGCAAAACGTTGCTGTACAGACTTGC	722
Oy	557	caatagacctgacgaaacggaaatgccaataaatgtgtgtatgtgttggaaanaaggt	616
Db	723	CAATCGTCTGTAGAGACCCGAATGTCAATAAAATGTGTGCTTGTCGAAACAGTGT	782
Oy	617	cgtgaagcgagttcaatgagtgtgtagctcccccgaagaatgtgtaactgcgtaactga	676
Db	783	GGTGAGCCAAATTAACGAGTGTGCGGTTTCCGAAAGAAATGTGTGCCCGGAAATTCG	842
Oy	677	tgtgtgtaacttctcctgtacccgagatcccgatgtctctgtccgaagattggaataaaga	736
Db	843	TGTGGGTGAATCCCGGTCGCGATGTAATGCAAGGTGTCAAAATTTTAAATGAAACA	902
Oy	737	tttaagcggaataatggtctactactcgagcttggaatcccaacttggatgctttagatg	796
Db	903	CTTATAGTGGAGTGGTATTAACAACTGGTTTAAATCCACATTGTGATGATTTGATGG	962
Oy	797	ccaattgatgagttccatacagaagaacaaactgtggtggaatttactcttgsgaat	856
Db	963	TCAACTTCATGAGTTTCATATGCAAAATGATTAACCTTGTTGGAACTTAACATGGCGCAT	1022
Oy	857	acgtaacctgtatggagatttttactgcatacagcggtgcgaataatctgtgaagaatcc	916
Db	1023	AAAACTTTGATGGTGTGTTTTTTACTCGATCTGCTGTGCAAAACATTTGTCAGATCC	1082
Oy	917	aaagtatccgggatactactcaatcatgatataatgatctctctccacaagaatgag	976
Db	1083	AGATCTTCTGGAGCACTTTATATCAAGCAAAATGAGTTTCTTCAACCAAGATGACTG	1142
Oy	977	gtataatttgcataccaagaatgaataaagccagagatcacatatttgytactaa	1036
Db	1143	GTACATATTATTTCTCCAAATCGAAAAGAAACAAACCGACGATTAACATTCGTATCTACCG	1202
Oy	1037	gggcagaataatgcatgagatgcatgatatgtgtgtctgtacttcaacaagaagtgcagt	1096
Db	1203	AGGTGCAAAACGACGTGAGTGTAGCGTGGGTCCGTGATTTTACACCCGAAGTCGAC	1262
Oy	1097	tttgtctgaagaatataatccggaggttgaagaacgcagctcaaaaagtgtggcgctgatt	1156
Db	1263	ACATCCCGGATGCACCATCCCAAAACCTTACAAAAGCAGCAAAATCCGTGGGTGCAACTT	1322
Oy	1157	caacacatcataaaacagacaatacatgtggtccctgaaacctccctctgttgagaagtt	1216
Db	1323	TAAACATTTTCATACAAACGCAAAATGTTGTGGGCTGTAGCTCCATTTGGTGGAAAGCT	1382
Oy	1217	ggaggaagaagtgtgaagaggaagaagcgatcataaagaagtgtgagagatagaaga	1276

or

Query Match	39.2%	Score 622.6	DB 1	Length 1555
Best Local Similarity	74.3%	Pred. No. 1,9e-154		
Matches 819	Conservative	0	Mismatches 269	Indels 15
				Gaps 2
QY 424	ttgtctcaaaagctgtagcggttgatgctctcaagacttctacttgctctcagaagatg	483		

```

Db 364 TTGTCATCTGCAGATGAGTGTGATGCACTTAAACCTTGTCATGCTTATTTGAAGGAT 423
Qy 484 gcaaggttagagctctcgaaagtgcaatctcgaaacctcgatgtaagtaattgctcgtc 543
Db 424 GCAGGATAGAACTCGCAAGATGCTGCAACCCCTGCTGTCACCACTATGCGGTGCC 483
Qy 544 tccagacttgcaacaaatagactcgaaagaaagaaatgcaataaataatggtgattgt 603
Db 484 TTGAGACCTGCAATACCTCCAGATGAACCGAGTGCACATTTAAATGTGGGATCTGT 543
Qy 604 tgaagaacagctcgttagaagagtgcaatgagtgagctcccgaaagaagaatgttac 663
Db 544 TTGAGAACAGTGTGTGATGAGTCAACGAGTGTGCTGCTGTCAGAAAAGTGTGTC 603
Qy 664 ctcgtaaatctgattgttgtaacttctgtaactgataccagtgctctgtccagaagt 723
Db 604 CTGAGAAATCTGATCTCGAGATTTCTCTGCCAGACCCCTTGCTGTGTACAGAACT 663
Qy 724 ttgacatgaagaatttagcgggaatggttcatactactcgggttgtaaccactttg 783
Db 664 TCACATCTCGACTTTAACGGGAAGTGTACATTTACATGCTTCAATCCACCTTTG 723
Qy 784 atgcttgatcgcaatgcatgagttccatacaaga---gaaacaacattgtgtgga 840
Db 724 ATGCTTGATCTGCGACTGCGAGTGTGATGATCCACAGAAAGTGAACAAAGCTTGTGGA 783
Qy 841 attactcttgagaataactgtaacactgtagagatttctactgtagaggtgtgcaaa 900
Db 784 ACATCTCTTGGAGATTAAGACCTACAGATGATTTCTTACTAGTACGCGTACAA 843
Qy 901 aatcgtgcaagatcccaaatgatalccggggaactctacaatacgaataatgtaactc 960
Db 844 AATTCGTGCAAGATCTTAAACCAACCTGTGTCTCTACATATGACAAAGATACCTTC 903
Qy 961 tctaccaagatgtaggtatatttgcataccaagtagaagaatgctcgaaggtttca 1020
Db 904 ACTATCAAGATGATGATGATATCTGTCATCAAGATTAAGATTAACCTGAGACTTA 963
Qy 1021 tattgtgacttaagggcagaatgtagatgtagatgtagatgtagatgtagatgtagat 1080
Db 964 TATTGTATCTACCTGCGGCAAGAGATGCTGCTGATGATGATGATGATGATGATGAT 1023
Qy 1081 acaagaagatgtaggttgcctgtaagcattataccggagttgtaaacgcgaactca 1140
Db 1024 ACACGAGAGATGCTGATTAACCAATAGCATTTATCCAGAACTCGAAAAACACCAAAA 1083
Qy 1141 aagtgtaggttattcaacacatcataaacaagacaatacatgtagcctgaactc 1200
Db 1084 GCATAGGCGAGAGATCTCAACCATTTATTAGAACCGATTAACATGTTGCTGTAACCTG 1143
Qy 1201 cccctgttgaaggttgaagaagaatggaagaagaagaagaagaagaagaagaaga 1260
Db 1144 CCGTGTGAGAGATGAGATTTGAAGACAGTGAAGAAAGTGAAGATTAATGTAAGAGG 1203
Qy 1261 ttgagagatagaagaagaatggaagaagt-gagagataaagaag-----tca 1308
Db 1204 TTGAAGAGATGAAGAAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1263
Qy 1309 ccttctgaataactgtttgaaggtttaaagagcccaagagatggaagaactct 1368
Db 1264 CTTGTTCCAGAGATGAGCTCAAGATTTAATGAACCTGAAGCAAGAGAGAGAGAGAG 1323
Qy 1369 taagagactgaagaagaagaagaatgtagtcttgaatgtagtaataatggaagaactg 1428
Db 1324 TGAAGAGATTTAAGTAAGAGAGATGAAGATTTTGTGATGATCAAAATGAAGCAAGTG 1383
Qy 1429 aggtagaagaactttggtggttcttaccataaagaagaatgtagtaattttaa 1488
Db 1384 AGGTTGAAGAAATGTTTGGGAAGCTTTGCCATTCAGAGAGGTCAAGTAGAAGCAAGAC 1443
Qy 1489 aacatacaatatactacat 1511
Db 1444 CACCATGTTGTACAAACTATAT 1466

```

```

RESULT 4
ID 087587
AC 087587 standard; DNA; 1686 BP.
DE 19-DEC-1995 (first entry)
DEF DNA encoding Leucocytozoan protozoa structural protein epitope.
KW Leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;
OS Leucocytozoan protozoa sp.
PN J07089995-A.
PD 04-APR-1995.
PF 10-SEP-1993; 226078.
PR 10-SEP-1993; JP-226078.
PA (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
FA (NITS-) NISSEIKEN KK.
DR WP1: 95-167252/22.
PT Immune inducing polypeptide against Leucocytozoan protozoa - useful
PT in production of vaccines for treatment of leucocytozoanosis in
PS Claim 1: Page 12-14; 20pp; Japanese.
CC 087587-89 encode polypeptides having a whole or partial epitope of a
CC structural protein of Leucocytozoan protozoa (see R70491-93). The
CC polypeptides and DNA encoding them are useful in the production of
SQ vaccines for the treatment of leucocytozoanosis of fowl.
Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T;

Query Match 3.3%; Score 52.4; DB 1; Length 1686;
Best Local Similarity 52.3%; Pred. No. 0.00017;
Matches 116; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 1218 gagaagaagatggaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1277
Db 226 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 285
Qy 1278 gaagtagaagaagtgagaataaagaagtcacattatcagtaactggttgaagtttt 1337
Db 286 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 345
Qy 1338 aagaagctccacagagatgagaagaactcttaagaagctgagcaagaagaagaatgat 1397
Db 346 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 405
Qy 1398 gtttgaatgacttaaatggaagaagaagaagaagaagaagaagaagaagaagaaga 1439
Db 406 GAAGATGAACAAATGAAGATGAACAAATGAAGAAAGAAAGAAAGAAAGAAAGAAAG 447

RESULT 5
ID T91902
AC T91902 standard; DNA; 4000 BP.
DE 07-MAY-1998 (first entry)
DEF Mannose-1-phosphate transferase gene MNNA.
KW Mannose-1-phosphate transfer; MNNA gene; enzyme; yeast; regulatory gene;
KW human; high mannose type neutral saccharide chain; ss.
OS Saccharomyces cerevisiae.
FH Key location/Qualifiers
FT 418..3953
FT /tag= a
FT /product= MNNA
PN J09266792-A.
PD 14-OCT-1997.
PF 29-MAR-1996; 075667.
PR 29-MAR-1996; JP-075667.
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
DR WP1: 97-553460/51.
DR P-PSDB; W30763.
PT Positive regulatory gene of mannose-1-phosphate transfer in yeast -
PT useful for high mannose type neutral saccharide chain production

```


100

1

QY	1281	gtgaagaagtgagaggttaagaagatccacttttcagtaaaccttttaaggttttaa	1340
Db	3979	ATATTAAAAGACTGTAAGAATAATCAAGAAGCTTGAAGTGAATTTTGTAGACATATTATAA	4038
QY	1341	gagctccacagagatgaagaagaactctcttaagaagctgaacaaagaagaaatggatgt	1400
Db	4039	GAATTAAAACTATTGGAACAGATATTTTGTAGAGAAAAAAGAAATAGAAAAATGATCAT	4098
QY	1401	ttggaatggaacttaaatatgagaagcaactgagatgataaaaacttttggagctgtcttaca	1460
Db	4099	TTTGAAAAATTCGAGAGAGAGAGCTGGAAGAAATTTAAAGATCTTGAAACACATTTATTAAAA	4158
QY	1461	ataagaagaatgaatgaatgatattttaaactatcaacatatactatcatatgcat	1515
Db	4159	GAACTATCTTCATTAGACAGTTGGAAGAGAAAAAAATTTGAGAGAGCTACACGAT	4213

Search completed: October 14, 1999, 01:53:07
Job time: 3377 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 13, 1999, 18:23:30 : Search time 1349.22 Seconds
(without alignments)
2323.087 Million cell updates/sec

Title: US-09-075-375-3
Perfect score: 1589
Sequence: 1 tattttcattgagttgcagt.....tcttttcagaagaacaaaaa 1589

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: em_est20: *
21: em_est21: *
22: em_est22: *
23: em_est23: *
24: em_est24: *
25: em_est25: *
26: em_est26: *
27: em_est27: *
28: em_est28: *
29: em_est29: *
30: em_est30: *
31: em_est31: *
32: em_est32: *
33: em_est33: *
34: em_est34: *
35: em_est35: *
36: em_est36: *
37: em_est37: *
38: em_est38: *
39: em_est39: *
40: em_est40: *
41: em_est41: *
42: em_est42: *
43: em_est43: *
44: em_est44: *
45: em_est45: *
46: em_est46: *
47: em_est47: *
48: em_est48: *
49: em_est49: *
50: em_est50: *
51: em_est51: *
52: em_est52: *
53: em_est53: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	423.2	26.6	584	47	AI483033	AI483033 EST242356
2	201	12.6	458	24	N37612	N37612 18839 Lambd
3	172	10.8	602	50	AI668224	AI668224 60501BD09
4	45.2	2.8	527	43	AU034306	AU034306 AU034306
5	44.6	2.8	251	47	AI511110	AI511110 UI-R-BPO-
6	44	2.8	221	45	AI345703	AI345703 lb84b07.x
7	43.6	2.7	371	34	AA465230	AA465230 aa24g03.r
8	43.6	2.7	302	35	AA572217	AA572217 v151e08.r
9	43	2.7	196	41	AI053983	AI053983 q165e03.x
10	42.8	2.7	222	45	AI344147	AI344147 lb88a04.x
11	42.8	2.7	435	38	AA776195	AA776195 ae80e09.s
12	42.8	2.7	610	51	AI723679	AI723679 BSM13524
13	41.4	2.6	214	45	AI335076	AI335076 ta44d05.x
14	41.2	2.6	156	41	AI053452	AI053452 q178f01.x
15	41	2.6	387	38	AA804625	AA804625 nv98a08.s
16	40.4	2.5	361	49	AI644452	AI644452 m170g05.x
17	40.2	2.5	337	26	W79711	W79711 zd86f02.r1
18	40.2	2.5	469	28	AA064890	AA064890 zms50c02.r
19	40.2	2.5	290	28	AA090795	AA090795 YY0062.se
20	40.2	2.5	259	31	AA328524	AA328524 EST32037
21	40.2	2.5	317	39	C83992	C83992 C83992 Dict
22	40	2.5	391	36	AA622645	AA622645 np22b10.s
23	40	2.5	587	37	AA675470	AA675470 vr72c06.s
24	40	2.5	522	37	AA675519	AA675519 vr72e06.s
25	40	2.5	121	41	AI053748	AI053748 q173e02.x
26	40	2.5	121	41	AI053822	AI053822 q174e04.x
27	40	2.5	574	43	AU034427	AU034427 AU034427
28	39.8	2.5	537	40	C92734	C92734 C92734 Dict
29	39.8	2.5	216	44	AI311041	AI311041 ta45f05.x
30	39.6	2.5	475	38	AA773580	AA773580 ab61d02.s
31	39.4	2.5	552	31	AA311157	AA311157 EST181947
32	39.4	2.5	360	41	AI003258	AI003258 an07c05.s
33	39.4	2.5	210	44	AI308406	AI308406 lb43e06.x
34	39.4	2.5	233	48	AI580601	AI580601 lb71h04.x
35	39.2	2.5	354	20	T36200	T36200 EST98342 Hu
36	39.2	2.5	561	29	AA167835	AA167835 zq40f02.r
37	39.2	2.5	591	48	AI568863	AI568863 to23c02.x
38	39.2	2.5	360	50	F33733	F33733 HSPD27357 H
39	39	2.5	257	25	W06466	W06466 T2065 MVAT4
40	39	2.5	373	27	W40685	W40685 mc66a07.r1
41	39	2.5	634	35	AA549907	AA549907 0980m3 gm
42	39	2.5	483	41	AI005976	AI005976 ua80h08.r
43	39	2.5	340	49	AU053226	AU053226 AU053226
44	39	2.5	373	50	AU061684	AU061684 AU061684
45	38.8	2.4	1020	49	AL048658	AL048658 DKF2P566B

ALIGNMENTS

RESULT 1
AI483033 584 bp mRNA
LOCUS EST242356 tomato shoot, Cornell Lycopersicon esculentum cDNA clone
DEFINITION CLEBRK10, mRNA sequence.
ACCESSION AI483033
NID 94386957
VERSION AI483033.1 GI:4386957

KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Asteridae; euasterids I; Solanales; Solanales; Solanaceae; Solanum;
Potatoe; Lycopersicon.
1 (bases 1 to 584)
REFERENCE
AUTHORS van der Hoeven, R.S., Matern, A.L., Vision, T., Chandra, I.,
Ronnig, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W.,
Fraser, C.M., Venter, J.C., Martin, G.B., Giovannoni, J.J. and
Tanksey, S.D.
TITLE Generation of ESTs from tomato shoot meristem
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189131.

FEATURES
source
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU.
Location/Qualifiers
1. 584
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/map="18DE"
/clone="CLEB8K10"
/clone_lib="tomato shoot, Cornell"
/tissue_type="shoot meristem"
/dev_stage="8 week old plants"
/lab_host="XLOLR"
/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI; CLEB
- Tomato Shoot Meristem EST Library. Oligo-dT primed cDNA
library made from tomato vegetative shoots including
meristems and small expanding leaves."

BASE COUNT 179 a 110 c 124 g 171 t

ORIGIN

Query Match 26.6%; Score 423.2; DB 47; Length 584;
Best Local Similarity 87.2%; Pred. No. 1.1e-93;
Matches 464; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1 tatttcagagatttcagcttggtgtaatacgggtgaagaatgctcttcctccatt 60
DB 53 TATTACATGAGGTGGAGTGGTGGTTCATTCATTAATGAATAATGGCGCTGCCCTCAAT 112

QY 61 caaatlcttgagccaaacatgaacatcaatattatgtgtgaagcttccgctg 120
DB 113 CAACACTTCTGTCATCATTCAGAGCCATCAATATGTAATGGATCAGAGCTTCAGAGCT 172

QY 121 ataaaggtttagctggtgttggaagattacttctgtgtatagtcgtagcaaaatt 180
DB 173 ATCAAGATTGCTGGGTAGAGCAGATTACTTGTGTAGTACTAGTCAAGAAATT 232

QY 181 gtccagagagagactagacttgcgaataatcctcctagaataatgctgtgtttgg 240
DB 233 GTTCCAGAGACAGATCACTACATCTTGCAGAAATCTTCTGAAATATGTTGTGTTGG 292

QY 241 attcaagagcttcgaacatctcacaaggaacacacatctctcccgacataga 300
DB 293 ATTCTAGAGTCTGCACATCATTCACGCGGGAACAAATTTCTTCTTGCACATAGAA 352

QY 301 ttaaccagaatgtacctaaggaataatcagagatgcgaatctccaaagaatgacttga 360
DB 353 TTAACCGAATGTACTTAAGGAAATACAAATATGGAATTTCCAGAAAGATGAGCTTGA 412

QY 361 tggtttggaagaatggcgcaatttgcgaagaagcaattgtactatattcatttgt 420
DB 413 TGCTCTTGAAGAAATGGGCAATTGGCCAAAGACGAATGTATATTATTGT 472

QY 421 cagtgcttcaaaagctgacggtgatgctcctcaagacttgacttgactgaag 480
DB 473 CAGTGGCTCAAAAGCGCGATGCCCTTATGCTCTGCACAAATTTGACTGTTACTGAAAG 532

QY 481 agtgcaggttagagcttgcgaagtgcatttcgaacccctgactatgcaactaa 532
DB 533 AGTGAGGATGAGCTTGCAGAGTGCATCAACCTGCATGTGCAGCTAA 584

RESULT 2
LOCUS N37612 438 bp mRNA EST 05-JAN-1998
DEFINITION 16839 Lambda-PRL2 Arabidopsis thaliana cDNA clone 207C2377, mRNA
sequence.
ACCESSION N37612
NID g1158754
VERSION N37612.1 GI:1158754
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 438)
REFERENCE
AUTHORS Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H.,
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,
Retzel, E. and Somerville, C.
TITLE Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729
COMMENT On May 8, 1995 this sequence version replaced gi:799424.

CONTACT: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 2233tch@pml.cl.msu.edu
Seq primer: T7 dye primer.
Location/Qualifiers

FEATURES
source
1. 438
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="207C2377"
/clone_lib="Lambda-PRL2"
/note="Vector: Lambda Z1P-tox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark-rose; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques). The vector is BRL's Lambda Z1P-tox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dT primed cDNA."

BASE COUNT 108 a 97 c 102 g 114 t 17 others

ORIGIN

Query Match 12.6%; Score 201; DB 24; Length 438;
Best Local Similarity 76.3%; Pred. No. 2.5e-39;
Matches 257; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

QY 486 aggttagagcttcggaagtgacatttcgaacctgcatgtgacgactaattgtcctgtctc 545
DB 14 AGGATGAACTCGCAAGTGACATTCGCAACCTGCTGTGACGCAATGTCCGCTGCTT 73

Oy	546	cagatttcgaacaataaggcccttcagcaaacggaatctcagataaattggtgattgctt	605
Db	74	CAGACTTCGAATAACCGCTCCAGATTAAACCGAGTCCGAGTTAAATGTGGGATGTGTTT	133
Oy	606	gaaacagtlgtcgtagacgaagttcaatgaqtlgtcagttctccgaagaatlygtacct	665
Db	134	GAGAACAGTGTGTGTGATGAGAGTTCAACACAGTGTCTGTGTCCGAGAAAAAGTGTCTTCT	193
Oy	666	cytaactcgtatggtgtgagactctcctcctactcgtatcccaagttctgtgccagaagttt	725
Db	194	AGAAATCTGATCTCGGAGAAATTTCCTGCCCCAGACCTTCTGTCTGTATCAGAACTTC	253
Oy	726	gacaclyaaagatlttaagcgggaatlygttcattactcgcggtllyaaalccacatttgat	785
Db	254	AACATCTCGGACTTTAAACGGGAGAGTGTAATACAAATGCGTTGAATNCAACCTTTGAT	313
Oy	786	gcttttga-ctgcgaattgcagatgttcatacaga	821
Db	314	GCTTTCGACTTCGCCANCTGCATGGTTCCTCCNACAGCA	350

RESULT	/3
LOCUS	A1668224/c
DEFINITION	A1668224 602 bp mRNA EST 14-MAY-1999 605018D09.xl 605 - Endosperm cDNA library from Schmidt lab Zea mays
ACCESSION	A1668224
NID	94827532
VERSION	A1668224.1 GI:4827532
KEYWORDS	EST,
SOURCE	Zea mays.
ORGANISM	Zea mays. Euarvota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea. 1 (bases 1 to 602)
REFERENCE	Walbot V
AUTHORS	Maize ESTs from various cDNA libraries sequenced at Stanford University (1999)
TITLE	Unpublished (1999)
JOURNAL	On May 18, 1998 this sequence version replaced gi:3137458.
COMMENT	

FEATURES	Source	BASE COUNT	ORIGIN
	1. .602	149 a	164 c
	Location/Qualifiers	117 g	172 t
	/organism="Zea mays"		
	/cultivar="Ohio43"		
	/db_xref="taxon:4577"		
	/clone_lib="605 - Endosperm cDNA library from Schmidt lab"		
	/tissue_type="pericarp, embryo, and endosperm"		
	/dev_stage="10 days post-pollination"		
	/lab_host="DH5(alpha)"		
	/note="Organ: Kernel; Vector: PAD-GAL4-2'; Site:1: EcoRI; Site:2: XhoI; kernel endosperm cDNA library from Schmidt lab"		

	Query Match	10.8%	Score 172	DB 50	length 602
Db	Best Local Similarity	66.4%	Pred. No. 3.3e-32		
	Matches 247	Conservative	0	Mismatches 125	Indels 0
					Gaps 0
QY	1106	aagcattatccgaggttcacaaacgcgacgtcaaaagtgtggcgtgatlttcaacacatt	1165		
	600	AACAAATAATACCGAGACTTGGAAAGACTCTCAAAAGACCGTAGGTGGGACTTCTGCAGCGTT	541		

QY	1166	caataaaacagcaataaacctgagccctgcgaaccctctcttbaaggtttgaagaaga	1225
Dd	540	CACTATGACCGAAMAACCTGTGCTCTGACCTCTCTTTGTGGAGAGATCGAGAAAC	481
QY	1226	agttgaagaagttgaagaagcatataaaagaattgagagatagaagaagaataga	1285
Dd	480	TGTGTGTGGAAGSGAAMACCATTCGTACGGAGGTGAGAGATCATCGAGAGAGTTGA	421
QY	1286	gaaggtgaagagataaagaagctaccttctttagtaaaacttttgaaggttttaagaagct	1345
Dd	420	GGACTGTGGAAMGAGAGAGGCTCACTCTTTCCAGAGCTGCGAAMGTTCTCATGGAGGT	361
QY	1346	ccaagagatgaagagaaacttcttaagaagctgaagcaaaagaagaatgtaatttttga	1405
Dd	360	GAACACAGATTTTCATCAACTTCTTCAGAGGGCTGAGCAAGAGAGATGAGACTGTTTGA	301
QY	1406	tgaacttaaatgtgaagcaactcgagtgtaaaaaacttttgggagtgctttaccataag	1465
Dd	300	TCACTTAACAAT1TGAAGCGACTGAAGTTGAGCAAGTCTTCAGCCGTCACTACCGTTGAG	241

RESULT	4	
LOCUS	AM034306	
DEFINITION	AM034306 527 bp mRNA EST 28-APR-1999	
ACCESSION	AM034306 Dictyostellium discoideum SL (H.Urushihara) Dictyostellium	
NID	dictyosdum cDNA clone SLc554, mRNA sequence.	
VERSION	AM034306	
KEYWORDS	G3799730	
SOURCE	EST.	
ORGANISM	Dictyostellium discoideum.	
REFERENCE	Dictyostellium discoideum.	
AUTHORS	Eukaryota; Dictyostellida; Dictyostelium.	
TITLE	1 (bases 1 to 527)	
	Mori,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,	
	Yoshino,T., Mittra,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H.,	
	Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.	
	The Dictyostellium developmental cDNA project: generation and	
	analysis of expressed sequence tags from the first-finger stage of	
	development	
JOURNAL	DNA RES. 5 (6), 335-340 (1998)	
MEDLINE	95156227	
COMMENT	On Jan 14, 1998 this sequence version replaced gi:1877648.	

Contact: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
 Email: d402huesakura.cc.tsukuba.ac.jp
 PROJECT = Dictyostelium discoideum cDNA project in Japan
 POLYA-NC: Location/Qualifiers
 1..527
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /map="8p21..3-p22"
 /clone="SLC554"
 /clone.lib="Dictyostelium discoideum SL (H. Urushihara)"
 /dev_stage="sing"
 267 a 67 c 92 g 100 t 1 others
 BASE COUNT
 ORIGIN

Query Match	2.8%;	Score 45.2;	DB 43;	Length 527;
Best Local Similarity	53.4%;	Pred. No. 0.35;		
Matches 95;	Conservative 0;	Mismatches 83;	Indels 0;	Gaps 0.

FEATURES
 source
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.jnl.gov) for further information.
 Possible reversed clone: polyt not found
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 430.
 Location/Qualifiers
 1. 435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="11: 737E12: 16"
 /clone="IMAGE:970504"
 /clone_lib="Stratagene schizo brain S11"
 /sex="male"
 /tissue_type="schizophrenic brain S-11 frontal lobe"
 /dev_stage="34 years old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: Bluescript SK-; Site_1: EcoRI; library
 constructed from S-11 frontal lobe, male, 34 years old,
 50% caucasian, 50% Aleutian. Schizophrenic suicide.
 Random primed into EcoRI site of ZAP II Vector. Mass
 excised. Custom library. Avg insert length 1.4kb.
 Material obtained by Johnston N., Torrey, E.F., Yolken R.
 and the Stanley Neuropathology Consortium - Analysis of
 RNAs from the Brains of individuals with psychiatric
 Diseases (unpublished) Stanley Neuropathology Laboratory.

BASE COUNT	153 a	Johns Hopkins School of Medicine, Baltimore MD."	79 c	92 g	111 t
ORIGIN					
Query Match	2.7%	Score 42.8;	DB 38;	Length 435;	
Best Local Similarity	52.9%;	Pred. No. 1.3;			
Matches 92;	Conservative 0;	Mismatches 82;	Indels 0;	Gaps 0;	
Qy 1384	aagaagaatgagatgtgttgatgagacttaaatggaagcaactgagtgaaacactt 1443				
Db 102	AAAAAAAAAGAGAGAGAGAGAGAGACATGGGATGCGAAGAAAAAACACGCGAGAAAAAAGA 161				
Qy 1444	ttgggctgtccttaccacaataagaaatlaagatlaattttaaacctlcaacatata 1503				
Db 162	GTGGCTTTCTTTGGAATTAAGTACACATAAGTTACTGTTGGGGGATCTATCAAAATTTAA 221				
Qy 1504	tactaacgtatagtgtgtattgattccttttgccctggaatgattgctttaca 1557				
Db 222	GACTAGGTCAAACCTTACTTTTATTTCATTTTGTTGTCAAAGATGGCTATTTC 275				
RESULT 12					
LOCUS	A1723679				
DEFINITION	A1723679 610 bp mRNA EST 11-JUN-1999				
ACCESSION	BSBML3524586SK Brugia malayi infective larva CDNA (SAR94WT-BML3)				
NID	A1723679				
KEYWORDS	g5042543				
ORGANISM	A1723679.1 GI:5042543				
REFERENCE	EST.				
AUTHORS	Brugia malayi.				
TITLE	Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;				
COMMENT	1 (bases 1 to 610)				
	Williams, S.A.				
	Unpublished (1995)				
	On Dec 20, 1995 this sequence version replaced gi:1133752.				

BASE COUNT	ORIGIN	3	a	56	c	6	g	91	t
	/dev_stage="adult" /lab_host="DH10B" /note="Organ: Ovary; Vector: PAMP; RNA made from papillary serous ovarian carcinoma, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 500 bp. Primary library, non-amplified."								

	Query March	2.6%	Score 41.2;	DB 41;	Length 156;
	Best Local Similarity	67.48;	Pred No. 2.6;	Mismatches	0;
	Matches 56;	Conservative	0;	Gaps	0
QY	1209 gagaagtttggaagaagaatgtaagaagaacgcatacaaaagttagag	1268			
Dd	128 GAGGAGGGAGGAAGAAGCAAAAGCAGACAGAAAAGAAAGAAAGAA	69			
QY	1269 atagaagaagaattagagaagttag	1294			
Dd	68 GAAGCAACAAGAGAGAGAGAG	43			

RESULT	15	
LOCUS	AA804825/c	
DEFINITION	AA804825 387 bp mRNA EST 18-FEB-1998	
ACCESSION	AA804825	
VERSION	92873801	
KEYWORDS	AA804825.1 GI:2873801	
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
TITLE	Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
JOURNAL	1 (bases 1 to 387)	
COMMENT	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap .	
	National Cancer Institute, Cancer Genome Anatomy Project (CCAP),	
	Tumor Gene Index	
	Unpublished (1997)	
	On Jan 14, 1998 this sequence version replaced gi:1796874.	

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Kitzman, Ph.D.
 .cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bto.llnl.gov/dbfp/image/image.html
 Insert Length: 621 Std Error: 0.00
 Seq Primer: -40m13 fwd. ET from Amersham
 High quality sequence crop: 352.
 Location/Qualifiers
 I. 367

```

/map="17"
/clone="IMAGE:1237814"
/clone_lib="NCI_CGAP_Prl6"
/sex="male"
/tissue_type="tumor"
/lab_host="DH10B"
/note="Organ: Prostate; Vector: pAMP10; mRNA made from
invasive prostate tumor cells, CDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on

```

```

      agarose gel, average insert size 600 bp. Library made by
      D. Krieman, NIH.*
BASE COUNT      17 a      134 c      29 g      207 t
ORIGIN

Query Match      2.6%; Score 41; DB 38; Length 387;
Best Local Similarity 63.9%; Pred. No. 3.5;
Matches 62; Conservative 0; Mismatches 35; Indels 0; Gaps 0.

QY 1209 gagaggttggagaagaagctggagaagagagaaagcattcataaagaagttgagag 1268
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 GAAAGGAGAGCAACAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 41
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1269 atagaagaagagctgagaagaaggtggagaataaagaag 1305
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 40 AAGGAAGAAGAATTACTGATGAGAGAGAGAGAGAGAGAGAG 4
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: October 13, 1999, 18:23:36
Job time: 2504 sec

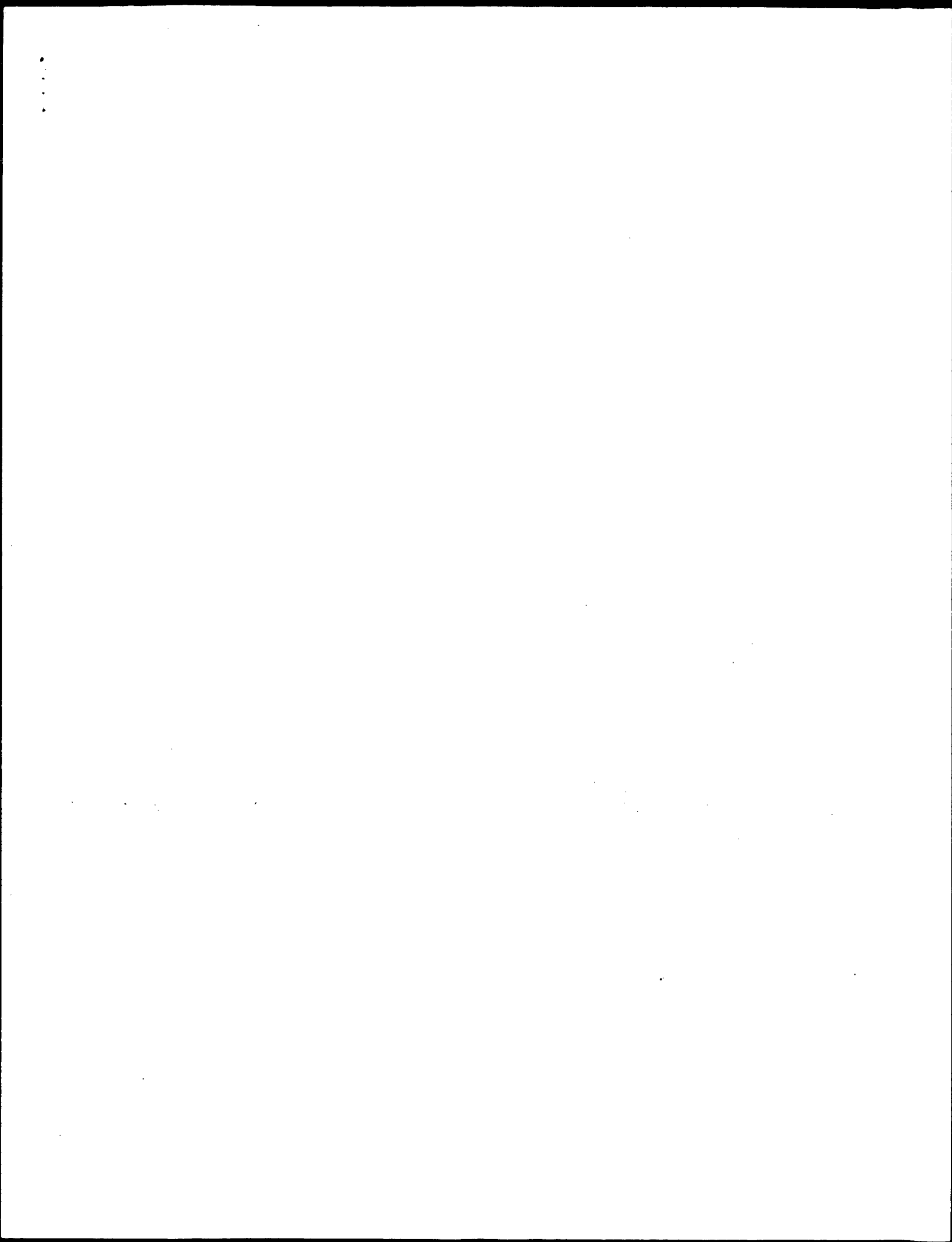
```

```
Search completed: October 13, 1999, 18:23:36
Job time: 2504 sec `
```

```

Qy 1209 gagaagttcgaaagaagaagtcgaaagaagaagcaccgatacataaagaagttgaggag 1268
    ||| | ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 100 GAAAGCAGCAGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 41
    ||| | ||||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Qy 1269 atagaagaagnagtcgagaaggtgagagataaagaag 1305
    | ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 40 AAGCAAGAAAGAACTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 1999, 03:56:10 ; Search time 34.13 seconds
(without alignments)
979.923 Million cell updates/sec

Title: US-09-075-375-4
Perfect score: 7495
Sequence: 1 MALSHVFLCKEALNDLYA.....MEASEVEKLFKALPIKRV 1412

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2554	34.1	478	1 W09875	Tobacco violaxanthin
2	2498	33.3	473	1 W09874	Romaine lettuce vi
3	2417	32.2	462	1 W09876	Arabidopsis violax
4	1475	2.0	1786	1 W24790	P. falciptarum live
5	147	2.0	2482	1 R72826	Human mitorin. Pur
6	147	2.0	2482	1 W23996	Human mitorin. Pur
7	135.5	1.8	1312	1 W71295	Human homologue of
8	135.5	1.8	1312	1 W22775	Human RAD50. Human
9	135	1.8	1639	1 W85011	p8alpha-green flo
10	134	1.8	1639	1 W54145	p8alpha-green flo
11	132	1.8	776	1 R60178	Lethal factor of B
12	129.5	1.7	2954	1 Y01632	Amino acid sequenc
13	129	1.7	1654	1 P50777	Sequence of the p1
14	129	1.7	993	1 R95268	Pre-nisin modifica
15	124	1.7	1582	1 R77087	Rat sulphonylurea
16	124	1.7	1498	1 R77084	Rat sulphonylurea
17	121.5	1.6	1132	1 R97866	Chicken leucocytos
18	121	1.6	1411	1 W02258	Nucleolar/endosoma
19	119.5	1.6	1604	1 R70105	TNF-R-EBA 175 fusi
20	119.5	1.6	1099	1 W40538	Mutant C-beta prot
21	119	1.6	1135	1 W31185	Human p160 polypep
22	119	1.6	905	1 W31186	Human p160 polypep
23	118.5	1.6	2273	1 R96811	Erysiphe graminis
24	115.5	1.5	1130	1 R71729	Merosin major subu
25	115.5	1.5	3110	1 R71730	Merosin major subu
26	115	1.5	1435	1 R70232	P. falciptarum SABP
27	115	1.5	1435	1 W22477	Sialic acid bindin
28	114.5	1.5	1230	1 W17785	Potato tuber solub
29	114.5	1.5	1743	1 W98795	H. pylori GHPD 175
30	112.5	1.5	3248	1 R99795	Kinetochore protei
31	112	1.5	649	1 W90005	Expressed antigen
32	111.5	1.5	1130	1 R13436	Merosin M polypept
33	111.5	1.5	2329	1 W25038	Partial BRCA2 cand
34	111.5	1.5	3418	1 W23287	Human breast and o
35	111.5	1.5	3418	1 Y04357	Human BRCA2 (om14)
36	111.5	1.5	3418	1 Y04354	Human BRCA2 (om11)
37	110.5	1.5	1786	1 R41043	CD4-EBAl175 fusion
38	110.5	1.5	1164	1 R85781	Group B Streptococ
39	110.5	1.5	1164	1 W40537	Group B Streptococ
40	110.5	1.5	1128	1 W40539	Mutant C-beta prot
41	110.5	1.5	1093	1 W40540	Mutant C-beta prot
42	110.5	1.5	1164	1 W40541	Mutant C-beta prot
43	110	1.5	751	1 W98340	H. pylori GHPD 325

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description
W09875	1	2554	34.1	478	1 W09875	Tobacco violaxanthin
W09875	2	2498	33.3	473	1 W09874	Romaine lettuce vi
W09875	3	2417	32.2	462	1 W09876	Arabidopsis violax
W09875	4	1475	2.0	1786	1 W24790	P. falciptarum live
W09875	5	147	2.0	2482	1 R72826	Human mitorin. Pur
W09875	6	147	2.0	2482	1 W23996	Human mitorin. Pur
W09875	7	135.5	1.8	1312	1 W71295	Human homologue of
W09875	8	135.5	1.8	1312	1 W22775	Human RAD50. Human
W09875	9	135	1.8	1639	1 W85011	p8alpha-green flo
W09875	10	134	1.8	1639	1 W54145	p8alpha-green flo
W09875	11	132	1.8	776	1 R60178	Lethal factor of B
W09875	12	129.5	1.7	2954	1 Y01632	Amino acid sequenc
W09875	13	129	1.7	1654	1 P50777	Sequence of the p1
W09875	14	129	1.7	993	1 R95268	Pre-nisin modifica
W09875	15	124	1.7	1582	1 R77087	Rat sulphonylurea
W09875	16	124	1.7	1498	1 R77084	Rat sulphonylurea
W09875	17	121.5	1.6	1132	1 R97866	Chicken leucocytos
W09875	18	121	1.6	1411	1 W02258	Nucleolar/endosoma
W09875	19	119.5	1.6	1604	1 R70105	TNF-R-EBA 175 fusi
W09875	20	119.5	1.6	1099	1 W40538	Mutant C-beta prot
W09875	21	119	1.6	1135	1 W31185	Human p160 polypep
W09875	22	119	1.6	905	1 W31186	Human p160 polypep
W09875	23	118.5	1.6	2273	1 R96811	Erysiphe graminis
W09875	24	115.5	1.5	1130	1 R71729	Merosin major subu
W09875	25	115.5	1.5	3110	1 R71730	Merosin major subu
W09875	26	115	1.5	1435	1 R70232	P. falciptarum SABP
W09875	27	115	1.5	1435	1 W22477	Sialic acid bindin
W09875	28	114.5	1.5	1230	1 W17785	Potato tuber solub
W09875	29	114.5	1.5	1743	1 W98795	H. pylori GHPD 175
W09875	30	112.5	1.5	3248	1 R99795	Kinetochore protei
W09875	31	112	1.5	649	1 W90005	Expressed antigen
W09875	32	111.5	1.5	1130	1 R13436	Merosin M polypept
W09875	33	111.5	1.5	2329	1 W25038	Partial BRCA2 cand
W09875	34	111.5	1.5	3418	1 W23287	Human breast and o
W09875	35	111.5	1.5	3418	1 Y04357	Human BRCA2 (om14)
W09875	36	111.5	1.5	3418	1 Y04354	Human BRCA2 (om11)
W09875	37	110.5	1.5	1786	1 R41043	CD4-EBAl175 fusion
W09875	38	110.5	1.5	1164	1 R85781	Group B Streptococ
W09875	39	110.5	1.5	1164	1 W40537	Group B Streptococ
W09875	40	110.5	1.5	1128	1 W40539	Mutant C-beta prot
W09875	41	110.5	1.5	1093	1 W40540	Mutant C-beta prot
W09875	42	110.5	1.5	1164	1 W40541	Mutant C-beta prot
W09875	43	110	1.5	751	1 W98340	H. pylori GHPD 325

CC (T66242): VDE nucleic acids (see also T66241, T66243), in sense
 CC or antisense orientation. Increased levels result in the plant
 CC modify VDE levels in plants. Increased levels result in the plant
 CC being tolerant of increased light and therefore more productive
 CC and/or more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 SO Sequence 478 AA:

Query Match 34.1%; Score 2554; DB 1; Length 478;
 Best Local Similarity 99.8%; Pred. No. 1,6e-193;
 Matches 476; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 474 MALAPSHNFIANHEITIKYVYSGKLPQHKKRPSMGMDYFGSIYAKICSSRRIPRYRKPSP 533
 DB 1 MALAPSHNFIANHEITIKYVYSGKLPQHKKRPSMGMDYFGSIYAKICSSRRIPRYRKPSP 60
 QY 534 RICCGDLSRGLQLFSGKHNLSPAHSINONVPKNSGCKFPKDVALMWEKNGOFAKTAI 593
 DB 61 RICCGDLSRGLQLFSGKHNLSPAHSINONVPKNSGCKFPKDVALMWEKNGOFAKTAI 120
 QY 594 VAIFIIISVASKADAVDALKTCTCLKECRLAKCISNPACAAVACLOTCNNRPDETGC 653
 DB 121 VAIFIIISVASKADAVDALKTCTCLKECRLAKCISNPACAAVACLOTCNNRPDETGC 180
 QY 654 QIKCGDLFENSVDNEFCVSRKCVPRKSDVDGFPVDPSPVLVCKPMPKPSGWFET 713
 DB 181 QIKCGDLFENSVDNEFCVSRKCVPRKSDVDGFPVDPSPVLVCKPMPKPSGWFET 240
 QY 714 RGLNPTFEDAFCOLHEFHFEENKLVGNLSWRIRTPDGGFFIRSAVQKEVQDPKPYGILYN 773
 DB 241 RGLNPTFEDAFCOLHEFHFEENKLVGNLSWRIRTPDGGFFIRSAVQKEVQDPKPYGILYN 300
 QY 774 HDNEYLLYQDDWIIISKVENSPEDYIFYYIKRNDAMGCGYGSVLYTSAVLPESIIE 833
 DB 301 HDNEYLLYQDDWIIISKVENSPEDYIFYYIKRNDAMGCGYGSVLYTSAVLPESIIE 360
 QY 834 LQTAAGKVGDFNTFTIKTDNTGCEPPLVERLEKKVEEGEERTTIKVEEIEEVEKVRK 893
 DB 361 LQTAAGKVGDFNTFTIKTDNTGCEPPLVERLEKKVEEGEERTTIKVEEIEEVEKVRK 420
 QY 894 EYTLFSKLEEGFKLQDEENFLRELKSEMDVLDGLKMEATEVEKFGRALPIRKL 950
 DB 421 EYTLFSKLEEGFKLQDEENFLRELKSEMDVLDGLKMEATEVEKFGRALPIRKL 477

RESULT 2

W09874 2
 ID W09874 standard; Protein: 473 AA.

AC W09874;
 DT 28-JUL-1997 (first entry)

DE Romaine lettuce violaxanthin de-epoxidase.
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW Photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 OS xanthophyll; lettuce.

OS Lactuca sativa L. cv. romaine.
 FH Key location/Qualifiers

FT peptide 1..125

FT protein /label= Transit_peptide

FT peptide /label= Mat_protein

FT domain /note= "Claim 8"

FT peptide /label= Cys-rich_domain

FT domain /label= Lipocalin_signature

FT domain /label= Highly-charged_domain

FT peptide /label= Tryptic_peptide-11

FT peptide 275..289

FT FT peptide /label= Tryptic_peptide-21
 FT 341..353
 FT /label= Tryptic_peptide-15
 FT misc_difference 132
 FT /note= "conserved Cys residue"
 FT misc_difference 134
 FT /note= "conserved Cys residue"
 FT misc_difference 139
 FT /note= "conserved Cys residue"
 FT misc_difference 146
 FT /note= "conserved Cys residue"
 FT misc_difference 152
 FT /note= "conserved Cys residue"
 FT misc_difference 158
 FT /note= "conserved Cys residue"
 FT misc_difference 162
 FT /note= "conserved Cys residue"
 FT misc_difference 171
 FT /note= "conserved Cys residue"
 FT misc_difference 175
 FT /note= "conserved Cys residue"
 FT misc_difference 190
 FT /note= "conserved Cys residue"
 FT misc_difference 197
 FT /note= "conserved Cys residue"
 FT misc_difference 243
 FT /note= "conserved Cys residue"
 FT misc_difference 373
 FT /note= "conserved Cys residue"
 PN W09717447-A2.
 PD 15-MAY-1997.
 PE 07-NOV-1996; US-018291.
 PR 07-NOV-1995; US-006315.
 PR 06-AUG-1996; US-023502.
 PA (CALJ) CALGENE INC.
 PI Bugos RC, Rockholm DC, Yamamoto HY;
 DR WPI: 97-281036/25.
 DR N-PSDB: T66241.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light.
 PS Example 1; Fig 1; 41pp; English.
 CC The 55 kDa violaxanthin de-epoxidase (VDE) (W09874) of romaine
 CC lettuce catalyses the de-epoxidation of violaxanthin to zeaxanthin
 CC and antheraxanthin. This system, termed energy dependent
 CC non-radiative energy dissipation or non-photochemical fluorescence
 CC quenching, reduces the quantum efficiency of photosystem II (PSII)
 CC helping to prevent PSII over-reduction and photoinhibitory damage.
 CC The amino acid sequence of the VDE was deduced from an isolated
 CC cDNA clone (T66241). VDE nucleic acids (see also T66242-43), in
 CC sense or antisense orientation, can be used in genetic constructs
 CC to modify VDE levels in plants. Increased levels result in the
 CC plant being tolerant of increased light and therefore more
 CC productive and/or more resistant to disease. Underexpression of
 CC VDE increases photosynthetic efficiency under low light. The
 CC photosensitivity of a range of crops, trees and ornamentals can be
 CC modified.
 SO Sequence 473 AA:

Query Match 33.3%; Score 2498; DB 1; Length 473;
 Best Local Similarity 99.6%; Pred. No. 4,1e-189;
 Matches 471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MALSLHIVFLCKREBALNVLARSPCNFRHSGOPPIITIMAKIRSNNGFNFRFLTSTYK 60
 DB 1 MALSLHIVFLCKREBALNVLARSPCNFRHSGOPPIITIMAKIRSNNGFNFRFLTSTYK 60
 QY 61 TTSFSDSHCKDKSQCSIDTSPFEEIQRPDLKRGMTLIEKQMPOTQAIYLVCFYIV 120
 DB 61 TTSFSDSHCKDKSQCSIDTSPFEEIQRPDLKRGMTLIEKQMPOTQAIYLVCFYIV 120
 QY 121 PRVDVADALTKACALLKEGRIELAKCIANPSCAANYACLOTCNNRPDETECOIKGDLFE 180

Db 121 PRVDALAKTACILKECRLEAKCIANPSCAANVACLOTCNNRPDETQIKCGDLFE 180
 Qy 181 NSVYQNECAVSRKKCKPRKSDGEPVPDPAVQVQNNKDSGKWTYISGLNPFEDA 240
 Db 181 NSVYQNECAVSRKKCKPRKSDGEPVPDPAVQVQNNKDSGKWTYISGLNPFEDA 240
 Qy 241 FDCQHEHFMENDKLVGMLTWIKTLDGCFTRSAVQTFVODPLPGALYNHNEFLHYQ 300
 Db 241 FDCQHEHFMENDKLVGMLTWIKTLDGCFTRSAVQTFVODPLPGALYNHNEFLHYQ 300
 Qy 301 DDWYILSSQIENKPPDYFVYVYRGNDAMDYGSGVYTRSPITPESTIPMLQKAASVG 360
 Db 301 DDWYILSSQIENKPPDYFVYVYRGNDAMDYGSGVYTRSPITPESTIPMLQKAASVG 360
 Qy 361 RDNFNFTTDSGCEPPLVERLEKTAEGEKILKEAVETEEVEKVEKVRDTEMTLF 420
 Db 361 RDNFNFTTDSGCEPPLVERLEKTAEGEKILKEAVETEEVEKVEKVRDTEMTLF 420
 Qy 421 QRLGFEKLEODEENFVRELSEKEKILNELQMEATEVEKLFGRALPIRKL 473
 Db 421 QRLGFEKLEODEENFVRELSEKEKILNELQMEATEVEKLFGRALPIRKL 473

RESULT 3
 ID W09876
 AC W09876 standard: Protein; 462 AA.
 DT 28-JUL-1997 (first entry)
 DE Arabidopsis violaxanthin de-epoxidase.
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll.
 OS Arabidopsis thaliana var. Columbia.
 FH Key Location/Qualifiers
 FT peptide 1..113
 FT /label=Transit_peptide
 FT protein 114..462
 FT /label=Mat_protein
 FT peptide 114..126
 FT /note="Claim 8"
 FT domain 114..185
 FT /label=Cys-rich_domain
 FT domain 364..462
 FT /label=Highly-charged_domain
 FT misc_difference 120
 FT /note="conserved Cys residue"
 FT misc_difference 122
 FT /note="conserved Cys residue"
 FT misc_difference 127
 FT /note="conserved Cys residue"
 FT misc_difference 134
 FT /note="conserved Cys residue"
 FT misc_difference 140
 FT /note="conserved Cys residue"
 FT misc_difference 146
 FT /note="conserved Cys residue"
 FT misc_difference 150
 FT /note="conserved Cys residue"
 FT misc_difference 159
 FT /note="conserved Cys residue"
 FT misc_difference 163
 FT /note="conserved Cys residue"
 FT misc_difference 178
 FT /note="conserved Cys residue"
 FT misc_difference 185
 FT /note="conserved Cys residue"
 FT misc_difference 231
 FT /note="conserved Cys residue"
 FT misc_difference 362
 FT /note="conserved Cys residue"
 PN W09717447-A2.
 PD 15-MAY-1997.
 PF 07-NOV-1996; U18291.

PR 07-NOV-1995; US-006315.
 PR 06-AUG-1996; US-023502.
 PA (CALJ) CALGENE INC.
 PI Bugos RC, Rockham DC, Yamamoto HY.
 DR WPI, 97-281036/25.
 DR N-PSDB; T66243.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 PS Disclosure: Fig 3: 4pp: English.
 CC The violaxanthin de-epoxidase (VDE) (W09876) of Arabidopsis
 CC catalyses the de-oxidation of violaxanthin to zeaxanthin and
 CC antheraxanthin. This system, termed energy dependent non-radiative
 CC energy dissipation or non-photochemical fluorescence quenching,
 CC reduces the quantum efficiency of photosystem II (PSII), helping to
 CC prevent PSII over-reduction and photoinhibitory damage. The amino
 CC acid sequence of the VDE was deduced from an isolated cDNA clone
 CC (T66243). VDE nucleic acids (see also T66241-42), in sense or
 CC antisense orientation, can be used in genetic constructs to modify
 CC VDE levels in plants. Increased levels result in the plant being
 CC tolerant of increased light and therefore more productive and/or
 CC more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 SQ Sequence 462 AA.

Query Match 32.2%; Score 2417; DB 1; Length 462;
 Best Local Similarity 99.8%; Pred No. 9.9e-183;
 Matches 461; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 951 MAAVATGFTSPCHDRIRFFSSDDIGRLGTRKRINGFLKILPLPGADRTTGRSS 1010
 Db 1 MAAVATGFTSPCHDRIRFFSSDDIGRLGTRKRINGFLKILPLPGADRTTGRSS 60
 Qy 1011 RPLSAFRSGSKGIFDIPLPSKNEIKELTAPLLKLVGLACAFIYVSAVDAKTC 1070
 Db 61 RPLSAFRSGSKGIFDIPLPSKNEIKELTAPLLKLVGLACAFIYVSAVDAKTC 120
 Qy 1071 ACLKGRILKACIANPACANVACLOTCNNRPDETQIKCGDLFENSVDNECAV 1130
 Db 121 ACLKGRILKACIANPACANVACLOTCNNRPDETQIKCGDLFENSVDNECAV 180
 Qy 1131 SRKKVPRKSDLGEPAPDSVLYQNFNISDFNGKWTYISGLNPFDAFCOLHEFHTEG 1190
 Db 181 SRKKVPRKSDLGEPAPDSVLYQNFNISDFNGKWTYISGLNPFDAFCOLHEFHTEG 240
 Qy 1191 DNKLVGNIWRKTLDSGFTTRSAVQTFVODPQVLYNHNEFLHYQDDWYILSSKTE 1250
 Db 241 DNKLVGNIWRKTLDSGFTTRSAVQTFVODPQVLYNHNEFLHYQDDWYILSSKTE 300
 Qy 1251 NKPEDYIFVYVYRGNDAMDYGGAVYTRSSVLPNSIIPLEKAAKSIRDFSTFRTDN 1310
 Db 301 NKPEDYIFVYVYRGNDAMDYGGAVYTRSSVLPNSIIPLEKAAKSIRDFSTFRTDN 360
 Qy 1311 TCGEPALVERIEKTYEGERITVKEVEEIEEVEKVEKVRTEMTLFLQRLAEGFNEIK 1370
 Db 361 TCGEPALVERIEKTYEGERITVKEVEEIEEVEKVEKVRTEMTLFLQRLAEGFNEIK 420
 Qy 1371 QDEENFVRELSEKEMFLDEIKMEASVEKLFGRALPIRKYR 1412
 Db 421 QDEENFVRELSEKEMFLDEIKMEASVEKLFGRALPIRKYR 462

RESULT 4
 ID W24790
 AC W24790 standard: Protein; 1786 AA.
 DT 08-OCT-1997 (first entry)
 DE P. falciparum liver stage antigen-3.
 KW Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
 KW prophylaxis; Thai strain; Gene organisation: exon; intron; hydrophobic;
 KW glycosyl-phosphatidylinositol; membrane anchoring sequence; antibody;
 KW vaccine; immunotherapy; malaria.

OS Plasmodium falciparum.
 FH key Location/Qualifiers
 FT region 223..278 /note="repeat region 1"
 FT region 279..818 /note="repeat region 2"
 FT region 1537..1578 /note="repeat region 3"
 PN WO9641877-A2.
 PD 27-DEC-1996.
 PF 12-JUN-1996; F00894.
 PR 13-JUN-1995; FR-007007.
 PA (INSP) INST PASTEUR.
 PI Daubersies P, Druilhe P;
 DR WPI: 97-055464/06.
 DR N-PSDB: T78867.
 PT Plasmodium falciparum poly:peptide(s) and related nucleic acids -
 PT derived from the liver stage antigen-3, useful for malaria vaccine
 PT prodn. and diagnosis
 PS Claim 1: Fig 2A-I; 69pp; French.
 CC This sequence corresponds to a Plasmodium falciparum strain K1
 CC pre-erythrocytic liver stage antigen-3 (LSA-3) protein. The encoding
 CC gene sequence was isolated by screening a P. falciparum strain T9/96
 CC library with serum from a missionary treated by prophylaxis (for strain
 CC T6/96 see FR9101286). Of 20 clones isolated, clone 7295 was used to
 CC screen a library generated from Thai strain K1. One clone contained a
 CC 6.85 kb insert including the genomic sequence T78867. The gene comprises
 CC a 1.8 kb region encoding 3 major blocks of tetrapeptide repeats
 CC (especially the amino acid sequence VEEs, VEEN, VEEI, VAPS, VAPT, etc)
 CC and a 3' hydrophobic region corresponding to a glycosyl-phosphatidyl-
 CC inositol membrane anchoring sequence). The invention relates to new
 CC polypeptides of at least 10 amino acids derived from the LSA-3 protein
 CC with the exception of the peptides W24791-4. The LSA-3 peptides can be
 CC used to raise antibodies and as vaccines for immunotherapy of malaria.
 SQ Sequence 1786 AA;

Query Match 2.0%; Score 147.5; DB 1; Length 1786;
 Best Local Similarity 18.1%; Pred. No. 0.013;
 Matches 207; Conservative 155; Mismatches 374; Indels 409; Gaps 50;

QY 347 SIIPNQR-AAKSVGRDN-----NFTTNSCGPEPLVE-----RLKTAEG 390
 DB 858 TLENEVETTESVTFEINLEIQTENTNDTI---EKLELHEENVLSALENTQSEE 914
 QY 391 EKLITKEAVELEEEVEKEV-----EKVADTMTLPQRLLEGFKLQD---EENYREL 441
 DB 915 EK---KEVIDVIEEKEVAVATLLEIYEQAEEKSANTITTEFENLEENAVESNENVAENL 971
 QY 442 SKEREIINLEOMEATEVEKLEGRALPIRLMALAPHSNFLAHNETIKRYVGSKLPGHK 501
 DB 972 EKLNETVNTLVKVEEVEIISGESLENEMDKAF-----FSEIFDNVKGIOELTLTG-- 1025
 QY 502 RFSMGWEYFSGSIYAKICSSRRILPRYFRKSPRICCGIDSGFLGFSHG---KHNLSPA 557
 DB 1025 -----MFSIETISITQSEE--KVDLNNVVSIIIDN--IENMEGLINKLENISST 1072
 QY 558 HSIQNVKPG-NSGCKFEPRKDVALVWKEKMGOFATAITVAIFILSVAKAAYDALKTCTC 616
 DB 1073 EGVQETVTEHEQVNVYDVDPANK-----DOF-----LGLINAGG----- 1110
 QY 617 LKACRELELAKCISNPACANVACLOTCNNRPDETECO---IKGDLFEKSVVDFENEC 673
 DB 1110 -LKEMFNLDEVEKFS---ESDVITVEIKDEPVQKEVEKEFVSIEMEMENIVDLEE-- 1164
 QY 674 VSRKKCVPRKSDVGD--FVPVDPVVLVOKFMDMDFSGKWTITKGLNFTPAFCQLHEFH 731
 DB 1164 -----EKEDLTDMKIDAVEESIEIS--DSKET-----EST 1193
 QY 732 TEENKLVGNLSWIRTPDGGFEFTSAVQKVFODPKRYPGILYNDHNEKLLVQDDWYLLSSK 791
 DB 1194 KDKREKDVSLVEEVQDND---MDESVEKYLE-----LKNHEELMKDAVINIDITSK 1242

QY 792 -VENSPEYIFVYKGRNDAMDGYGSLYTRSAVLPEST--IPLEOTAAQKVRGRDNFTF 848
 DB 1243 LIETQE-----LNEVEADLIKMEKLEKRA---LSDSKEI 1278
 QY 849 IKTDNTCGPEPLVERLEKVEEGE--RTIIEVEEI---EEVEKYVD-----K 893
 DB 1279 IDAKD-----DTLEKVEIEEHIDITTLDEVELKQVEEDKLEKSDKLDLEDILK 1329
 QY 894 EV---TLFSKLEFGKEIORDENFLRELKSEMDVDLGMETVEVKLEGRAL----- 946
 DB 1330 EYKEIELSEILEDYKELKTITETDLLEKKEIEKHDFEKEFEEDKLEADILKEVS 1389
 QY 946 -----PIRKLMAVAHTCFTSPCHDIRFRFSSDGIQRLITRKIRINGTFLKIL 994
 DB 1390 SLEVEEKKLEEVHELKEVEHIIISGDH-----IKG----- 1422
 QY 995 PPIQSADLTGTGSSRPLSARSGFSKGFIVPLPSKNELKEITADLLKLVGLACA 1054
 DB 1422 --LEEDLEEVDLKGSILDMKGMELGDMD-----KESLEDVTTKL----- 1463
 QY 1055 FLIVPSADAVDALKTCACILKGCRIELAKCIANPACANVACLOTCN--NRPDETECOIK 1112
 DB 1463 -----GERVESLKDVLSALGMDEDMKTRKAQRKLEVLK 1501
 QY 1113 CGDLFENSVDENECVARRKCVPRKSDLGEPAPDPSVLVQNFNISDFNGKWTISGL 1172
 DB 1502 -----EEVKEEPRKKIRKK-----VRDIDK-----K 1524
 QY 1173 NPTEPFDQLHEFHTEGDNKLVGNISRIKTLDSGFTTRSAVOKFVODDPQPGVLYNHD 1232
 DB 1525 EPRDELVEVEMKQDIED-----VEEDLEED----- 1552
 QY 1233 NEXLYODWYLLSSKIKENKPEYIFVYRGRNDAMDYGAVVYTRSSVLPNSIPELE 1292
 DB 1552 -----LEEDKVEDIDEDIDEDIGEDKDEVID-----LIYQKE 1583
 QY 1293 KAAKSGRDFSFITDNTCGPEPALVERIEKYVEEGRIIVKEVE-----IEEVE 1345
 DB 1584 KRIEYV-----KAKKKKLEKVEEGVSGIKKHVDEMKYVQKIDREVD 1626
 QY 1346 KEVEKY--GRTEMTLFORAEGFENLKODEE-----NFVELKKEMEPFDELKMEASE 1397
 DB 1627 KEVSKALEKKNVT-----NYLKONODFPKVKNFYAKKYVFAAPFISAAVAFASY 1677
 QY 1398 VEKLF 1402
 DB 1678 VVGFF 1682

RESULT 5
 ID R72826 standard; Protein; 2482 AA.
 AC R72826;
 DT 27-FEB-1996 (first entry)
 DE Human mitosis.
 KW Cell cycle; M phase; mitosis; retinoblastome; mitosis; cell growth;
 OS Homo sapiens.
 FH key location/Qualifiers
 FT region 1480..1659
 FT /label=internal_repeat
 FT region 1660..1839
 FT /label=internal_repeat
 PN WO9511309-A2.
 PD 27-APR-1995.
 PF 24-OCT-1994; U12162.
 PR 22-OCT-1993; US-141239.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Lee W, Zhu X;
 DR WPI: 95-170229/22.
 DR N-PSDB: Q86851.
 PT Purified mammalian protein mitosis and agents that bind it and
 inhibit its action - used to promote cell growth or to inhibit cell

PT division and/or proliferation.
 PS Claim 4; Fig 8B; 61pp; English.
 CC R72829 is human mitotin. Mitotin is involved in the regulation of
 CC the mammalian mitotic cell cycle. Mitotin as with E2f-1 (see R72824)
 CC interacts with the retinoblastoma protein (the retinoblastoma tumor
 CC suppressor gene product). Mitotin is first synthesized at the G1/S
 CC boundary, it is then phosphorylated from S through M phase, and during
 CC mitosis, is closely associated with the centromeres/kinetochores at the
 CC mitotic spindle poles. Mitotin is necessary for a eukaryotic cell to
 CC enter the M phase of the mitotic cell cycle and its degradation is
 CC necessary for a cell to advance on to the next stage. Mitotin is thus
 CC useful for controlling cell growth as overexpression of mitotin prevents
 CC a cell from exiting the M phase.
 CC An anti-mitotin antibody, antibody fragment or a phosphorylated mitotin
 CC mutin (or nucleic acid encoding it) can also be used to inhibit cell
 CC division which is particularly useful for the study of the cell cycle.
 CC A further use is to control hyperproliferative cells, and so control
 CC diseases such as psoriasis and breast cancer. It can also be used to
 CC block gametogenesis of an immature gamete.
 SO Sequence 2482 AA;

Query Match 2.0%; Score 147; DB 1; Length 2482;
 Best Local Similarity 18.4%; Pred. No. 0.024;
 Matches 240; Conservative 188; Mismatches 443; Indels 424; Gaps 60;

QY 379 LVRLEKTAIEGEGKLIKRAVEIEEVE-----KEYE-KVRDTEMTPQLLEGKELOQ 432
 DB 315 LSTSLLEKEMSSITSLNKRLEIELTQENGTLKEINASLNOKMLIOK-SESPANYID 373
 QY 433 DEENFYRELS--KEKEKIL----- 450
 DB 374 ERKESISELDQYKQKILLLOKCEFTGNAYEDLSQYKAQKNSKLECLINECTSICE 433
 QY 450 ---NEIOMATEVEKLFGRALP--IRKLMLALPHSNFLANHTTYGYGSKLPGH---- 501
 DB 434 NKRNELE---OLKEAFKHEHOFELTKLFAEERNONMLLELYOQALREXTDQNNNS 489
 QY 501 KRRSMGDEYFGSIYVAKICSSRRIPRYFKSPRI-----CCGLDSRGLOLFSHGK 551
 DB 490 KSPAGGIKQEIIMTLKEQNKNQKQEVNDILOENQOLMKVAKTRKECNLSEPR----- 544
 QY 552 HNLSPAHSINQNVKPGNSGCKFPKDYALMWEKGOFAITAIFILSVASKADAVDAL 611
 DB 544 -----NSVARESERNQ-CNFKPQMDLEVKETSLDSYNAQLVQLEPAMELNKELKQESE 596
 QY 612 KTCTCL-----LKECRLEL-AQCISNP-----ACAN 637
 DB 597 KKEKCIQHELOITRGDLETSNLODMOSQELSGLKDEIDEKYEYISGPHLESTQNDNAH 656
 QY 638 VAC-LQTCNNRPETE--COIKCGDLEFNSVDEPNCAVSRKCYA--PRK--SDGDF- 690
 DB 657 LQCSLQTTAKNLKEKICELQAEKE--LVLELND--SRSECTIATPKMAEYKGLL 711
 QY 690 ---PVDPSPVLVQKEDMKDFSGKWPITRGINP-----TFDAPDC 725
 DB 712 NEVKIILNDGGLHGLVEDIDPGEF--GEOPNEQHPSLAPLDESNSYEHLTLSDKEV 768
 QY 726 QLHEFHTEENKLVGNLSWRIRTPDGGFTFSANQKVFQDPKYGILYHNHEYLLOYD-- 784
 DB 769 QMRAELQEK-----FLSLQSEHKILHDOHQ--MSSKMSLQTYVDSL 810
 QY 784 --DWYIISKVENSPEDYIFVYKGRND-----AMDYGGSVLYTRSAVL 826
 DB 811 KAEMLVISTNRNFOGLVEMQGLEGLVPSLSSCVDSSLSLGGSSSY--RALL 868
 QY 827 PE---SIIPLOTAO---KYGRDFTNTFKTDNCTGPEPL-----VERLEKYEVEE 873
 DB 869 EOTGDMSTLSNLGSAVANOCSDVEFCSSLOEENLTRKEPPAPAKVGELESLSEYVR 928
 QY 874 RTIKKEVEIEEVEYK---VADKEVTLFSKLFEG-----FRELORDENLRELSE 922
 DB 929 QSL-----EKLEKMSOGIMKNKEIQLLEPLLSSEROELDLCKROYLSENOQMKLTSV 984

QY 923 EMDVLQDKMEATEVERKLF-----GRALP-IRKLMAVATH-----CFTS--- 961
 DB 985 TLEMSKLAIEKKQOTDLSLELVARLOGLDLSRSRLGIDTEADIAIGRNECDSISE 1044
 QY 961 -----PCHDRIRPFSSDDGIGRLGTRKRINFTFLIKIIPPI-----QSADL--- 1003
 DB 1045 HTSETRTERFKHD-VHOICDKQAQDQDLNDIEKITEGALK--PTGECSGEOSPNTYE 1100
 QY 1003 -----RTTGGRS-----SRPLSAFRSGSKGIEDIVLPSPK----- 1034
 DB 1101 PGEDEKTOGSGEFTISELSEFGPALVPMDLNGQEDIHNLQLRVKTSDENLRHLVIED 1160
 QY 1034 -----NELKELTAPLLKLVGLACAFILVPSADAVDALTCACLLKGRLEIARC 1084
 DB 1161 RDRKVESLWKNMELDKLHLOEQVQMT-----KIAC-----ILEKI 1199
 QY 1085 IANPACANVACIQTCCNRPDETECQIKCDLFEANSVDEPNCAVSRKCYVRKSDLG- 1144
 DB 1200 V-----GEL-----KR--ENSDLSE 1212
 QY 1144 --EPPADPSYIVQNFNISDENGKWTITSGINPTFPAFDCQLEHFEHTE---GDNKLVGN 1197
 DB 1213 KLETFSCDHQELQORVETSE-----GLNS-----DEMAHDSRREDIDGNVAKYN 1258
 QY 1198 ISWRIKTIDSGFTFRSAVQKFEVQDPNPQGVLYNHNDNYLH---YQDDWYILSSKIENK 1252
 DB 1259 DSWKERFLD-----VENLSIRSEKASIEHEALYLEADLEAVQTEKLCLEKNENK 1310
 QY 1253 P-----EDYIFVYRGRNDAMDGYGAVVYTRSSVLPNSIIELEKAASIGRDSFTI 1306
 DB 1311 QKVIVCLEEELSVTSESNOL--RGELDTMSKKTALDOLSEKMEKETOLESHOSECL 1367
 QY 1307 RTDNTGCPPEPAIVRIKRIEVEGERIIVKEVEIEEVE---VEKEVEKGRTEMTLFQRLA 1363
 DB 1368 HClOYAEAEVASKTELLQTLSSDVSELKDKTHQELQSLDQSLITKCELENQIA 1427
 QY 1364 EGFNELKODEENFVRE-----LSKEMEFLDEIK-MEASEVEK 1400
 DB 1428 ---QNKKEKELVKESESIOARLESSESDYEXLNSKALEALAEK 1468

RESULT 6
 W23996
 ID W23996 standard; Protein; 2482 AA.
 AC W23996;
 DT 28-MAY-1998 (first entry)
 DE Human mitotin amino acid sequence.
 KW Mitotin; phosphotubulin; mitotic cell cycle; antibody; analogue;
 KW inhibition; M phase; Antagonist; hyperproliferative cell; cancer;
 OS Leukemia; lymphoma; chromosome segregation.
 KM Homo sapiens.
 FH Key
 FT Domain Location/Qualifiers
 FT 258..280
 FT /note="leucine heptad repeat"
 FT Domain 340..362
 FT 564..593
 FT Domain 1387..1443
 FT 1885..1962
 FT Domain 2146..2188
 FT 2165..2187
 FT /note="leucine heptad repeat"
 FT Domain 2188
 FT /note="Bipartite targeting motif"
 FT /note="optionally C or G"
 FT Misc_difference 2189
 FT Misc_difference 2301
 FT Misc_difference 2303
 FT /label="Bipartite targeting motif"
 FT /label="Bipartite targeting motif"
 FT /note="optionally A or T"
 PN US5710022-A.
 PD 20-JAN-1998.

PF 24-OCT-1994; 328254.
 PR 24-OCT-1994; US-328254.
 PR 22-OCT-1993; US-141239.
 PA (TEXA) UNTV TEXAS SYSTEM.
 PI Lee W, Zhu X;
 DR WPI; 98-109817/10.
 N-PSDB; V09076.
 PT New isolated mitotin protein and gene - useful for, e.g. developing
 PT products for therapy and diagnosis of hyper-proliferative disorders
 PS Claim 1. Column 40-52; 43pp. English.
 CC This is the amino acid sequence for mitotin, a phosphoprotein
 CC necessary for the cell to enter mitosis. The protein's degradation is
 CC also necessary for the cell to advance into the next stages of mitosis.
 CC The mitotin protein, can be used to control the growth of cells. An
 CC anti-mitotin antibody, a mutant or a non-functional analogue of mitotin
 CC can inhibit the mitotic cell cycle by preventing the cells from entering
 CC the M phase, and over expression of mitotin or its functional
 CC equivalent, would inhibit the cycle by preventing cells from leaving the
 CC M phase. Antagonists to this protein can be used to control
 CC hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease,
 CC psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast
 CC cancer, sarcoma and other neoplasms, bladder cancer, colon cancer,
 CC lung cancer and various leukemias and lymphomas). Reintroduction or
 CC supplementation of lost mitotin function by introduction of the protein
 CC or nucleic acid encoding the protein into a cell can restore defective
 CC chromosome segregation, which is a marker of progressing malignancy.
 CC Malignant proliferation of cells can then be halted. The protein
 CC can also be used for the detection and diagnosis of hyperproliferative
 CC cells.
 SQ Sequence 2482 AA;

Query Match 2.0%; Score 147; DB 1; Length 2482;
 Best Local Similarity 18.4%; Pred. No. 0.024;
 Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 IVERLEKTAEESEKLLKEAVEIEEVE---KEVE-KVRDTEMLTFOBLEGFEKLEQ 432
 DB 315 LSETLSLEKKESSITSLNKRFEIELTQENGLKEINSLNQEKNMLIQ-SESPANID 373
 QY 433 DEENVEELS---KEEKEL----- 450
 DB 374 EEKESITSELDQYKOEKLLILOCEETGNAYEDLSQYKAAQEKNSKLECLNCTSLCE 433
 QY 450 ---NELQWATEVEKLFGRALP--IRKLMLAPHSNFIANHEITIKYVGSKLPGH--- 501
 DB 434 NKKNLE---QLKEAFKHEQELTKLAFAPERNQNLMLLELITVQALRSEMTDNQNS 489
 QY 501 KRFSGWEDYRGSIVAAICSSRIPIRFKSPRI-----CCGLDSRGLQLFSGHK 551
 DB 490 KSEAGGLQOEIMTLKEQNKQKEVNDLLOENQOLMKYMKTRHEQNLSEPIR----- 544
 QY 552 HNLSPAHISINQVPGKNSGCKFPKDVALMVEKMGQFAKTAIVAFILTSVASKADAYDAL 611
 DB 544 ---NSVKRESERBNQ-CNFKPOMLEVAKEISIDSTNAOLYDLEALRNKELKLOESE 596
 QY 612 KTCTCL-----LKECRLEL-AKCI SNP-----ACAN 637
 DB 597 KEKELOHELOTIRODELTSMLOMOQSEISGLKDCLEDAEKYISCPHELSTQNNAH 656
 QY 638 VAC-LQTNKNNPDETE--CQIKCGDLFENSVYDEPNECANSRKCV--PRK--SDVGF- 690
 DB 657 LQCSLQTTMNLNLEKICETILQAEKYE--LVTEIND--SRSCITATRKMAEVEKTL 711
 QY 690 ---PVDPPIVYOKFDMKDFSGKWFTRGLN-----TFDFDC 725
 DB 712 NEVILINDSGLLIGELVEDIPGGEF---GEQPNQHPVSLAPIDESNSTEHLISDKEV 768
 QY 726 QLHEHTEENKLVNLSWRIPTPGGFEFTRSAVOKFVQDPKPIILNHNNEYLLYOD-- 784
 DB 769 OMHFAEIOEK-----FLSLQSEKHLIHDOHCQ--MSKMSSELOTIVYDSL 810

QY 784 --DWYIISSKVENSPEDYIIVYKGRND-----AMDYGGSVLYTRSAVL 826
 DB 811 KAENLYLSTNLRNFOGLVEMOLGLEGLVPISLSSCVDPSSSLSDSSPY--RALI 868
 QY 827 PE---SIPELOTAQ---KYGRDNFIFIKDNTCGPEPL-----VERLEKKEE 873
 DB 869 EQIDMSLNLNLSGAVSANOCSDYEVFCSSLOEENLTRRETPAPAKGVLEELSLCEYVR 928
 QY 874 RTIIEVEEIEEVEK---VRDKEVTLFSKLEFG-----FKELODEENFLRLSKE 922
 DB 929 QSL-----EKLEEKMESQGIKNKEIQLBQLLSSEHQELDLCKRYLSNEMQOKITV 984
 QY 923 EMDVLOLKNHAEVEKLF-----GRALPIRKIMAVATH-----CPTS-- 961
 DB 985 TLEMSKLAEEKOTQOLSLLEVARLOGLDLSRSLSLIGDTEBALQGRNESCDSKE 1044
 QY 961 -----PCHDRIRFFSSDDIGRLITRRKINGTFLKILPPI-----QSADL-- 1003
 DB 1045 HTSETTERTKHD-VQIQCDKAQDNLNDIEKITETGALK--PTGECGSHQSPDTNVE 1100
 QY 1003 ---RTTGRS-----SRPLAFRSQFSKGIPIVLPSPK----- 1034
 DB 1101 PGEDRTQGSSECISELSESGPMALVPMDFLQNOEDIHNLQVRKETSNNRLHLHYED 1160
 QY 1034 -----NEKELTAPLLKLYGVLAFLVBSADAVDLKTCACILKCGRIELAKC 1084
 DB 1161 RDKRVESLLEKEDLSKHLLEVOULT-----KIEAC-----TELKRI 1199
 QY 1085 IANPACANVACLOTNNRPDETECIKCGDLFENSVDENECASVRKRCVPSDGL- 1144
 DB 1200 V-----GEL-----KK--ENSDLSL 1212
 QY 1144 --EPAPDPSVLYQNNIDENGKWTITSGLNPTDAPDCQHEHTE---GDNKLVGN 1197
 DB 1213 KLEIFYSCDQELLQRYETSE-----GLNS-----DLEMHAKSREDIGDVAKYN 1258
 QY 1198 ISMRITLDSGFETRSAYOKFVQDPNQPGLVYNHNEYLN-----YODWYIISKIENK 1252
 DB 1259 DSKWERFLD-----VNEISRIKSEKASLEHALLYUADLEVYQTEKLCLEKDNENK 1310
 QY 1253 P-----EDYIIVYRGRNDADWDYGAVYTRSSVLPNSIIPLEKAKSIGRDFTFI 1306
 DB 1311 QKVIYCLEEELEISVTSERNQL--RGELDTMSKKTALDQSEKKKEKTEQELSHQSCSL 1367
 QY 1307 RDNQTCGPEPALVERIEKIVEEGERIIVKEVEIEEF---VEKEVEKGRFEMTLFOLIA 1363
 DB 1368 HCLQVAEAVKREKTELLOTLSSDVSELKDKTHLOEKLOSLKESQALSLTRCELENOIA 1427
 QY 1364 EGFNELKODEENFVE-----LSKEMEFLIDEIK-NEASEVEK 1400
 DB 1428 ---QNKREKELLVSESLQARLSDESYEKLNVSKALDAIYER 1468

RESULT 7
 W71295
 ID W71295 standard; Protein; 1312 AA.
 AC W71295;
 DT 25-NOV-1998 (first entry)
 DE Human homologue of yeast RAD50.
 KW Human homologue; yeast RAD50; Drosophila Septin-2; Acyl-CoA synthetase;
 OS Immunomodulatory activity; identification; activated T-cell.
 PN WO9838306-A1.
 PD 03-SEP-1998
 PF 27-FEB-1997; U03159.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Dolganov G;
 DR WPI; 98-481207/41.
 N-PSDB; V59979.
 PT Novel human immunomodulatory poly(peptide)s - have homology to the
 PT yeast RAD50 or Drosophila Septin-2 proteins
 PS Disclosure; Pages 136-140; 15pp; English.

CC The present sequence represents a human homologue of the yeast
CC 5. cerevisiae gene RAD50. The present sequence has 35% overall
CC homology to the yeast RAD50 gene, and is expressed in activated
CC T-cells, testis, foetal liver and heart tissues. The specification
CC also describes sequences encoding human homologues of the
CC yeast RAD50, the Drosophila Septin-2 and Acyl-CoA synthetase. The
CC proteins have immunomodulatory activity. The nucleic acids and
CC population can be used to identify activated T-cells in a sample
CC encoding other proteins or other compounds having immunomodulatory
CC activity.
SQ Sequence 1312 AA.

Query Match 1.8%; Score 135.5; DB 1; Length 1312;
Best Local Similarity 19.7%; Pred. No. 0.073;
Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

QY 382 RLEKTAEEGE---KLIIKAVIEEEVEKEVKRDTEMTLFORLLEGGFKELOQDEENF 437
DB 398 KLVREROGEAKTANQIMNDFAKETLKQKQIDEIRD-KKTGLRIIEKSEI----- 450
QY 438 VRELSEKEKEELNELOMATEVEKLFGRALPIRKLMAALAPHSNFL-----ANHETIKY 491
DB 450 ---LSKKQ---NELKNVKEYELOQLEGGSDRIILELODELKAEELSKAKNSNVEITLKM 502
QY 492 YVGSKLPGHRRFSWGEDYFGSIYVAKICSSRRIPRFRRKSPRICGLDSRGLOLFSHGK 551
DB 503 EVIS-LQNEK-----ADLDRTLRK-----LDQEMQL-NH- 531
QY 552 HNLSPAHSINQNPKNKSGCKFPKDVALLMWEKGGFAKTAIYAFITLSVASKADAVDAL 611
DB 531 -----HTTTRIOMEMLTDKADKDEQIRKIKSRHSDLSLGLYF-----PNKKQLEMDL 580
QY 612 KTCGCLKECELELAKCISNPACANVACIQTGNNRPDEIECOIK-----CGDLFE 662
DB 581 HSKSKEINQTRDLAKL--NKEILASSQNNKHINNELKREEQLSSEYEDKLPDVCQSDF 638
QY 663 NSVYDEFNECAVSRKKCVPRKSDVGDPPVPSVLQKFMKDFSGKWFITRGINPTFDA 722
DB 639 ESDLDRLKE-----EIEKSSKRAMLAGATAYSQ 668
QY 723 FDCQLEHFEHTENKLVGNLSWRIRTPDGGFFTRSAVQKFEVQDPKYPGILYNHNEYLLQ 782
DB 669 FITQL---TDENOSCCPVCQRY-----FQTEAELOEIVSD----- 701
QY 783 DDMYILSSKVENSPEDYIFV--YKGRNDAMDYGGSVLYTRSAV-LPESIIPELQTAQ 839
DB 701 ---LQSKLRAPDKLKSTESLKKERKRDMLGLVPMROSIIIDLKEKEIPELKNKLQ 755
QY 840 KVGGRDFTFIKTD-----NTGCEPPLVERLEKKEVEGEERTIIE-- 880
DB 756 NVNRDIQR-LKNDIEBQETLLGTIMPEESAKYCLDTVTIMERFQMLNDVKEKIAQQA 814
QY 880 -----VEEIEEVEKVRKDEVTLFSKLEFGFKELQDEENF-----LRELKSEE 923
DB 815 KLOGIDLRITVOQVNOKEOKHKLDTVSSKIELNRKLIQDOOEOLOHKLSTTNELKSEK 874
QY 924 MDVLDGLKME-----ATEVEKLE 941
DB 875 LQISTNLQRRQOLEBQTVELSTEVSQSLY 902

RESULT 8
ID W22775
AC W22775 standard; Protein: 1312 AA.
DT 21-DEC-1998 (first entry)
DE Human RAD50.
KW Human; RAD50; DNA repair; tumour suppression; cancer; Septin-2;
OS Homo sapiens.
PN W09727284-A2.

PD 31-JUL-1997.
PF 24-JAN-1997; U01299.
PR 17-JUL-1996; US-687080.
PR 26-JAN-1996; US-592126.
PA (GENE-) GENELABS TECHNOLOGIES INC.
PI Dolganov G;
DR WPI; 97-393672/36.
DR N-PSDB; 775237.
PT Human tumour suppressor gene RAD50 - useful to detect
PT predisposition to, decrease risk of and treat cancer, also Septin-2
PT homologues
PS Claim 5; Page 82-86; 195pp; English.
CC The human RAD50 (RAD50) is involved in DNA repair and has tumour
CC suppression activity, can be used to detect predisposition to, decrease
CC the risk of or treat cancers, e.g. acute myeloid leukaemia,
CC myelodysplastic syndrome, therapy related myelodysplastic syndrome,
CC therapy related acute myeloid leukaemia, refractory anaemia or refractory
CC anaemia with excess blasts. Also disclosed in this invention is Human
CC Septin-2 homologues of which may be used as targets for cancer therapies
CC and central nervous system directed treatment methods, and to measure the
CC proliferative potential of selected cell types.
SQ Sequence 1312 AA.

Query Match 1.8%; Score 135.5; DB 1; Length 1312;
Best Local Similarity 19.7%; Pred. No. 0.073;
Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

QY 382 RLEKTAEEGE---KLIIKAVIEEEVEKEVKRDTEMTLFORLLEGGFKELOQDEENF 437
DB 398 KLVREROGEAKTANQIMNDFAKETLKQKQIDEIRD-KKTGLRIIEKSEI----- 450
QY 438 VRELSEKEKEELNELOMATEVEKLFGRALPIRKLMAALAPHSNFL-----ANHETIKY 491
DB 450 ---LSKKQ---NELKNVKEYELOQLEGGSDRIILELODELKAEELSKAKNSNVEITLKM 502
QY 492 YVGSKLPGHRRFSWGEDYFGSIYVAKICSSRRIPRFRRKSPRICGLDSRGLOLFSHGK 551
DB 503 EVIS-LQNEK-----ADLDRTLRK-----LDQEMQL-NH- 531
QY 552 HNLSPAHSINQNPKNKSGCKFPKDVALLMWEKGGFAKTAIYAFITLSVASKADAVDAL 611
DB 531 -----HTTTRIOMEMLTDKADKDEQIRKIKSRHSDLSLGLYF-----PNKKQLEMDL 580
QY 612 KTCGCLKECELELAKCISNPACANVACIQTGNNRPDEIECOIK-----CGDLFE 662
DB 581 HSKSKEINQTRDLAKL--NKEILASSQNNKHINNELKREEQLSSEYEDKLPDVCQSDF 638
QY 663 NSVYDEFNECAVSRKKCVPRKSDVGDPPVPSVLQKFMKDFSGKWFITRGINPTFDA 722
DB 639 ESDLDRLKE-----EIEKSSKRAMLAGATAYSQ 668
QY 723 FDCQLEHFEHTENKLVGNLSWRIRTPDGGFFTRSAVQKFEVQDPKYPGILYNHNEYLLQ 782
DB 669 FITQL---TDENOSCCPVCQRY-----FQTEAELOEIVSD----- 701
QY 783 DDMYILSSKVENSPEDYIFV--YKGRNDAMDYGGSVLYTRSAV-LPESIIPELQTAQ 839
DB 701 ---LQSKLRAPDKLKSTESLKKERKRDMLGLVPMROSIIIDLKEKEIPELKNKLQ 755
QY 840 KVGGRDFTFIKTD-----NTGCEPPLVERLEKKEVEGEERTIIE-- 880
DB 756 NVNRDIQR-LKNDIEBQETLLGTIMPEESAKYCLDTVTIMERFQMLNDVKEKIAQQA 814
QY 880 -----VEEIEEVEKVRKDEVTLFSKLEFGFKELQDEENF-----LRELKSEE 923
DB 815 KLOGIDLRITVOQVNOKEOKHKLDTVSSKIELNRKLIQDOOEOLOHKLSTTNELKSEK 874
QY 924 MDVLDGLKME-----ATEVEKLE 941
DB 875 LQISTNLQRRQOLEBQTVELSTEVSQSLY 902

RESULT 9
 ID W85011 standard; Protein: 968 AA.
 AC W85011;
 DT 08-FEB-1999 (first entry)
 DE p85alpha-green fluorescent protein fusion product.
 KW Human: p85alpha gene; fusion protein; green fluorescent protein; GFP;
 intracellular signalling; chimera.
 OS Chimeric - Aequorea victoria.
 PN WO9845704-A2.
 PD 15-OCT-1998.
 PF 07-APR-1998; DK0145.
 PR 07-APR-1997; DK-000392.
 PA (NOVO) NOVO-NORDISK AS.
 PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,
 PI Tullin S;
 DR N-PSDB: V71026.
 PT Determining effect on signalling pathways in live cells from
 PT redistribution of luminophores - specifically fusions of green
 PT fluorescent protein with a signalling component, and new apparatus,
 PT particularly for identifying toxins and potential therapeutic agents
 PS Example 15; Pages 96-98; 326pp; English.
 CC The present sequence represents a human p85alpha-green fluorescent
 CC protein fusion product. The fusion protein is used in an assay
 CC that exemplifies the invention. The specification describes how
 CC qualitative information about the influence of a molecule on a cellular
 CC response is obtained by recording the variation, caused by the molecule,
 CC on mechanically intact living cells, in the spatially distributed light
 CC emitted from a luminophore present in the cells. The variation in light
 CC emission is processed to provide information that correlates spatial
 CC distribution to the degree of the molecule. The method is used to
 CC identify agents that (in)directly affect intracellular signalling,
 CC especially to screen for potential therapeutic agents or toxins, and
 CC to identify new drug targets.
 SQ Sequence 968 AA;

Query Match 1.8%; Score 135; DB 1; Length 968;
 Best Local Similarity 19.8%; Pred. No. 0.05;
 Matches 180; Conservative 111; Mismatches 298; Indels 318; Gaps 49;

QY 687 GDFPVPDPVSL-----VQKFD-----MKDFSGKWFTRGLMPTDAPDCOLHEPTEEN 735
 DB 52 GLPLPMPPLVLTLLTYGVOCSFRYPDHMKQ-----HDFKSAPEGIVQER 97
 QY 736 KLVGNLSMRIRTPDGGFFTRSAVOKFVODP-----KYPGILYNHDEYLLYQD 783
 DB 98 TI-----FFKDDGNYKTRAEV-KFEGDLVNRLELKGIDKEGDNILGHKLEYNVNSH 149
 QY 784 DWYILSSKVENSPEDYIFYYKGRNDAMDYGGSVL-----YTRSAVDEES--IIELEOTA 837
 DB 150 NYVIADKKKNG---IKYNEKIRHNIEB---GSVOLADHYQNPIDGPIVLPD----- 199
 QY 838 AQKVGDFEFTFKIDNTCGPEPILVERLEKVEGEERTIKVEFEIEEVEKVRKEVTL 897
 DB 199 -----NHVLTOSALSADPN-----EKRDHNVL-----LEVTAAGITL 232
 QY 898 FSKLEFGEFELQDEENFLRELISKEEMD--VLDDIKMEATEVEKLFGRALPIRKLMAVA 954
 DB 233 -----GMDELTK-----SGLRMSAGEYQYRALYDYKKEHEEDIDLHGDITLVNKGSLVA 283
 QY 955 THGFTSPCHDRIRFFSSDDGIGRLGTRKRINGTL-----LKILPPIQASDLRTGGRS 1009
 DB 284 IG-FSDGQEARPEELGWLNGYNETTGEGRDEPRTYVEYIGRKRISPP-----TPKPRP 335
 QY 1010 SRPLSAFRGFSKGIPLDI-----VPLPSKNEL---KELTAPILKLKLVG-----LACAPLI 1057
 DB 336 PRPLPV-ARGSSKTEADVEQOALITPLDLAEQFAPDIAPIPLIKLVEALEKKGLECSITLY 394
 QY 1058 -----VPSAD-----AVDALKTACALLKGCRIE----- 1081

DB 395 RTQSSNNLAEINQILDCCDTPPSVDLEMDIVHLADAFKRYILDLPVPIPAAYSEMISLA 454
 QY 1081 -----LACIANPAC-----AANVACLOTCNNR----- 1104
 DB 455 PEVQSEEVYIQLKLRIRSPSIPHQYWLTLQYLKHFEKLSOTSSKNLNAVLSEISFP 514
 QY 1104 -----PDEECQIKCGDGLFENSVDNEFCNCAVSRKKVCPRRSDJGFPAPPSVLY 1154
 DB 515 MLFRFSAASDNTENLTK---VIEIITSTEMNEROPA--PALPP-----PPKPTTVA 562
 QY 1155 ---QNFENISDFNGKMY-----ITSGINPTFDA-----FDCOLHEFT-----EGDN 1192
 DB 563 NNGMNNMNSLOANEMWGDGISREVENEKLRDADGFLVROASTMHGDTTLTKGNN 622
 QY 1193 KLVGNLSMRIRKTIIDS--GR--FTNSAVOKFVOD--PNQGVLYNH--DNEYTH----YQD 1240
 DB 623 KLI-----KIFHRDGKYGSDPLTFSSVVELINHYRNESLAQYNRKLIVKLYPYSKYQO 677
 QY 1241 DWYILSKIE---NKPEDYIFYYRGRNDAMDYGGAIVYTRSSVLPNSIPELEKAKS 1297
 DB 678 DQYVKEDNIEAVGKLAHEY-----NT---QFOEKSR 706
 QY 1298 IGRDFFSTFRTDNTCGPEPALVERLEKTV-----EEGRITVEVEIEEBE-VEKEVE 1349
 DB 707 YDRLYEYRTSQEIQMKRTAIEAFNETIKIFEDQOQDERYSKYEIEKREKNEKEIQ 766
 QY 1350 KVRTEMTLEFORLAEFNELKODENFVRELSEKEMEFIDEIKMASEYEKLFQALP-- 1408
 DB 767 RI-----MHNVDKLKRSISEIIDSRRLEEDLKQAAEY-----REIDKRMNSIKBDL 814
 QY 1408 --IRKVR 1412
 DB 815 IQLRKTR 821

RESULT 10
 ID W54145
 W54145 standard; Protein: 1639 AA.

AC W54145;
 DT 23-SEP-1998 (first entry)
 DE P. falciparum synthetic gp190 protein.
 KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 KM monoclonal antibody; passive immunisation; parasite.
 OS Plasmodium falciparum.
 PN Synthetic.
 PD WO9814583-A2.
 PF 09-APR-1998.
 PD 02-OCT-1997; E05441.
 PA 02-OCT-1996; DE-040817.
 PR (BURA/) BUIARD H.
 PI Buiard H, Pan W, Tolle R;
 DR WPI: 98-240088/21. V35363.
 DR N-PSDB: V21451. V35363.
 PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
 PT protein - useful in anti-malaria vaccines, also stabilising genes by
 PS reducing their AT content
 PS Example 1; Fig 3c; 48pp; German.
 CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface) protein. The gene encoding this protein has been
 CC stabilised by reducing the AT content of the nucleotide sequence. Such a
 CC protein is useful in vaccines against malaria or for producing monoclonal
 CC antibodies (for passive immunisation). The complete gp190 protein can now
 CC be produced outside the parasite and has, at least over extended regions,
 CC the native pattern of folding. Larger amounts of the protein can be
 CC produced recombinantly than would be possible using the parasites as
 CC source.
 SQ Sequence 1639 AA;

Query Match 1.8%; Score 134; DB 1; Length 1639;
 Best Local Similarity 18.0%; Pred. No. 0.14;
 Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

```

QY 184 VQFNECAVSRKCYPRKSDVGEFPPDRNAVQNFNMKDESGKVIITSGLPTFPAFC 243
DB 249 IENINELLIEESKRTIDK-----NKNATKEEKKKLQAOYDLS-----LYNK 290
QY 244 QHEHEMENDKLVNLTMRKIKTLDDGFFTSRAVOF--VQDP-----DLPGALYNHNE 295
DB 291 QLEFAH---NLISVLEKRIIDLKKNENIKELDLKINIKNPANGNPNTLLDKNKK 346
QY 296 FLHYODWYILSSQJENKPPDY-----IYYGRNDAMDYGGSVYITTSPLPESII 349
DB 347 IEHEKEIKEIKTIFKNIDSFTDPLEYLEYLRKKNKNI--ISAKVETKESTEPNE-Y 403
QY 350 PNIQRAKASVGDPNFNIITDNGCPPEPLVERLEKTAEEGEKLLI-----KEAVEIEEE 404
DB 404 PNCVTPPLSY-NDINNALWELNSFG---DLNPFDPYTEPSKNITTDNEKRFINEIKER 459
QY 405 VEKEVEKVDTEMTLFFQRLLEGFKELQODEENFVRELSKEKEILNEL-----QMEAT 457
DB 460 IKIEKKKI-----ESDKSYEDRSKSLNDITREYKELNLEYDSKFPNNIDLT 507
QY 458 EYEKLFGR--ALPIKRLNALAPHSNPLANHETIKYYGSKLPGRKRSWGMEDY-EGSI 514
DB 508 NEKMGKGRSYVEKTL-----THNHTFASYENSKHNL-EKLTAKLY--MEDYSIRNI 558
QY 515 VYAKICSSRRIPRYERKSPRICGGLDSRGOLFSGKHNLSPAHINQNPKNNGSGC-- 572
DB 559 VVEKEL-----KYTK-----NLISKIENIT--ELTVENIKKDEQOLFEX 595
QY 572 -----KFPDVALWMEKMGQFAKTAIYAIFILSVASKADAVDALKTCTCLKECRLE- 625
DB 596 KITKDEKKPEKILEV-----SDIVKQVOVKLLM-----NKIDELKKTOLLKKNVEKLE 645
QY 625 -----LAKCSN-----PACAA-----NVACLOQCNRRPET 651
DB 646 NIHVPNSYKQENKQEPYILVILKEIDKLKVPYKESLINEKKNKITEGOSDSEST 705
QY 652 EECOI-----KCGDLFENSVDENECASRRKCYPRKSDVGEFPPDPVSVLVQ-- 700
DB 706 EGETGATATTKPQOQASALEG---DSVQAQAOEQKQAP-----PVPVPEAKQV 756
QY 700 -----KEDMDFSGSKWFTITKGLNTPFAPCOJLHEHTEENKLVGLNSMRI 745
DB 757 TTPAPVNNKTENYSKIDYLE-----KLYEFL-----NTSYIC 788
QY 746 RTPGGGFTSAVQKFPQDPKYPGILYNHD--NEYLLYODDWYILSSKYEN-----SPED 798
DB 789 H-----KY--ILVSHSTNMEKILKO--YKITKEESKLSQCDPLD 824
QY 799 YIF-----VYYKGRNDAMDYGGSV-----LYTRSAVL-----PESIIPELOTAQ 839
DB 825 LLFMIQNNIPMY-----SMFDSLNNLSQJFMEIYEKEMVCNLKDKDKIKNLEAK 880
QY 840 KVGEDNFTIKTIDTCGPEP-PLVERLEKKEVEGERT-----IIEVEEL----- 884
DB 881 KY-----SYVKTLLSSSMQPLSTLPQDKPEVSANDITSHSTNLSNKLFEENILSGNKK 936
QY 884 -----EEVEKRVKDEKVTLSKLFEGEKELQODEENFVRELSKEEEDYLDGK 931
DB 937 NIYQELIQKSSSENFYEKILKQSDTFYNESFTNFVSKADINDSLNDESKRK-----K 989
QY 932 MEATEVERKLFGRALPIRKLAVALATHCFTSPCHRIREFSSDDGIGRLGTRKINGITFLL 991
DB 990 LE-EDINKL-----KKTLOLSPDLNKKYKLELRFDKKKTIVGKYMOKIKLTL---LL 1038
QY 992 KILPFIQASDLRTIGGRSRP--LSAFRSGFSKIGIDYIPPLSKNKLKLTAPLLKLIV 1048
DB 1039 K-----EOLSKLNSLNPKHVLQNFVFFNK-----KKAFA----- 1071
QY 1049 GVLACAFIIVSADAVDALKTACCLLKGR--IELAKCIANPACAAVACLOTCNNRPDE 1106
DB 1071 -----IAETENLENTKILKIKHYKGLVYKNGESSPLKTISEESIQEDENVASL 1119
QY 1107 TECQIKCGDLFENSVDENECASRRKCYPRKSDVGEFPPDRNAVQNFNMKDESGK 1166

```

```

DB 1120 E--NFKVLSKILGKIKDNLN---LEKKLLSYLSSGLHHLLA-ELKEVINKKNY----- 1167
QY 1167 YITSSGLNTPFAPPCQJLHEPTEGDKNLVGNISMRKIKTIDSGFFTSRAVQKFPQD----- 1222
DB 1167 --TGNSPS-----ENFTDVNNALE---STK-KFLPEGIDVATVVSQSSDILEOS 1210
QY 1222 -PNOGVLVYNHNEFLHYQDDWYILSSKIKENKPEDYIFYYYGRNDA-WDYGGAVALYTR 1279
DB 1211 QPKRA-----SHVGAESNTITTSQNVQDEVDVITVPIFSESEDDYDGLQGV--TG 1262
QY 1280 SSVLPNSIIFPEIKAKSIGRDEFTIRIDNTGCPPEPALVERLEKTAEEGEKLLI----- 1339
DB 1263 EAVTPSVI-----DN-----ILSKIE--NEYEVLKPLAG 1291
QY 1340 IEHEVEKEVEKVGRTMTLFFQRLLEGFKELQODEENFVRELSKEKEFLD 1389
DB 1292 VYRSLLKOLE--NNVMTFNVAKDILNSRPNKRNKKNVLESLLIYKD 1338

```

RESULT 11

```

ID R60178 standard: Protein; 776 AA.
AC R60178;
DE 03-APR-1995 (first entry)
DT Lethal factor of Bacillus anthracis.
KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
  protective antigen; cell killing; targeting; pathogen;
  intracellular; HIV; human immunodeficiency virus; toxin.
OS Bacillus anthracis.
PN WC9418332-A.
PD 18-AUG-1994.
PE 14-FEB-1994; U0624.
PR 12-FEB-1993; US-021601.
PS 25-JUN-1993; US-082849.
PA (USSH N) US DEPT HEALTH & HUMAN SERVICES.
PI Arora N, Kimpel K, Lepia SH, Nichols PJ, Singh Y;
  MPI: 94-279753/34.
DR N-PSDB: Q70179.
PT Nucleic acid encoding anthrax toxin fusion protein - useful for
  targeting toxin to specific cells, eg for killing tumour cells
  or HIV-infected cells.
PS Disclosure: Page 75-77, 124pp; English.
CC The sequence encoding the lethal factor of Bacillus anthracis may be
  used in the construction of a nucleic acid which encodes a fusion
  of protein comprising the anthrax protective antigen binding domain of
  the native anthrax lethal factor and a sequence encoding an activity
  CC inducing domain of a second protein. The fusion proteins are useful
  CC for the specific killing of tumour cells or the killing of cells
  CC infected with intracellular pathogens, especially HIV.
SQ Sequence 776 AA.

```

Query Match 1.8%; Score 132; DB 1; Length 776;

Best Local Similarity 17.2%; Pred No. 0.061; Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

```

QY 733 EENKLVNLSWARTPDGFFTSRAVQKFPQDPKYPGILYNHD--NEYLLYQDDWY--I 787
DB 72 EMYKALIG--KIYVDGDTKHTISLSELSDEKKIKIDYGDALDHEHYVAKGYBPV 128
QY 788 LSSKVENSPEDYFVYIYKGRNDAMDYGGSVLYTRSAVLPESIIPELOTAQKQVARDNT 847
DB 129 L--VIOSSDDYVENTERKALN-----VYIEGKILSRDILSKINOPYOKFLDVANT 176
QY 848 FIKTDNCG-----PEPLVERLEKKEVEGERTIITK----- 879
DB 177 IKNASDSDGQDLFTNOLKRPFDVSEVFLQNSNVQVFAKAKAYIIEPQHRVQLY 236
QY 879 -----EVEEIEVEKRVKDEVTLSKLFEGKRE----- 908
DB 237 APEAFNYMDKFNQEIINLSLEELKDQRMLSREKMEKIKOHYQHSWDSLSEGRGLKKL 296

```

QY 908 ---LQDENFELRELSEKEMVDLGLKMEATEY---EKLFGALPI-----RKL 950
 Db 297 QIPLEPKDDILHLSQEEKELKRIQDSDSLTEEFKELKRIQDIDRSLSEEEKL 356
 QY 951 MAVAHCTFPCDRIRFFSSDGIIRLGTTRKRINGFLIKLIPPIQASD---LRTTG 1006
 Db 357 L-----NRLOVSSNP-----LSEK--EKEFLKRLKIDIPDYINQRLQDTG 396
 QY 1007 GRSSRPL-----SAFRSGFSKGIPIVPLPSKELKELTRAPLLKLVG 1049
 Db 397 GLDPSINLDVKKYKRIQIDALLHOSISSTLYNKLYLENNMINLATL----- 451
 QY 1050 VLACAPLIPADAVDALKTACILKGRICELAKIANPACANVACLOTGNNRDETEC 1109
 Db 451 -----GADLVDSST----- 459
 QY 1110 QIKCGDLFENSVDY---FNECAVSRKKCYPRK---SDIGEPAPDPVLYONENISDFNG 1164
 Db 459 -----DNTKINRGIFNEFKNFYKYSISSNYMTVDINERPADNERLKWRIQLSPDR 510
 QY 1165 KWTITGLNPTFPAFCQLEHFEHTEGDKLV--GNISRIKTLDSGFFTRBAVQKPVQDP 1222
 Db 511 AGYLENG-----KLILQRIKIGLEIKDVQ----- 534
 QY 1223 NQPGVLYNHNDNEYLAHODDWYLLSSKIENKRPEDYLFVYRGRNDAMDYGGAVVYTR--S 1280
 Db 534 ---IKQSEKEKIRI--DAKVVPKSKIDTKIQE---AQLNINQMNKALGLPKYTKLIT 584
 QY 1281 SVLPNGLIPELEKAKSISGRSTFIRTDNCGPPALVERLEKTEVEGE--RIIVKEY-- 1338
 Db 585 FNVHNNYASNIYSAVALLINENKNNIQSD-----LIKVTYLVQNGNRFVETDITL 636
 QY 1338 -----EEIEVEKEVEKVGRTMTLFORLAEGFNLKODEENFVRE----- 1380
 Db 637 PNIAEQYTHQDEIYEQVHSGKLYPESRSILHGPSKV--ELRNDSGFIHEHNAVDDY 695
 QY 1380 ---LSKEEME-----FLDEIKMEASEV 1398
 Db 696 AGYLLDKNOSDLVNSKFFIDIFKEGSGNL 725

RESULT 12
 ID Y01632 standard; Protein; 2954 AA.
 AC Y01632;
 DT 22-JUN-1999 (first entry)
 DE Amino acid sequence of centromere-associated protein-E (CENP-E).
 KW CENP-E, centromere-associated protein-E; Alpase activity;
 KW plus end-directed microtubule motor activity; chromosome congression;
 KW microtubule binding activity; chromosome movement; mitosis;
 KW cell proliferation; tumor; metastasis; vascular malfunction;
 KW inflammatory disease; immune disease; angiogenesis; hypertension;
 KW restenosis; fungal infection; selective herbicide; fungicide;
 KW insecticide; plant growth regulator; activator; cancer cell marker.
 OS Xenopus sp.
 PN WO9913061.A1.
 PD 18-MAR-1999.
 PD 10-SEP-1998: U19231.
 PR 11-SEP-1997: US-058645.
 PA (RBCG) UNITV CALIFORNIA.
 PI Cleveland DW, Goldstein LSB, Sakowicz R, Wood KW,
 DR N-PSDB: X26819.
 PT Centromere-associated protein-E and related nucleic acid
 PS Claim 5: Page 66-67; 77pp. English.
 CC The present sequence represents CENP-E (centromere-associated protein-E)
 CC of Xenopus. The protein has at least one of plus end-directed microtubule
 CC motor activity, Arpase (adenosine triphosphatase) activity and
 CC microtubule binding activity. CENP-E is the motor that powers chromosome
 CC movement toward microtubule plus ends and is essential for congression
 CC of chromosomes during mitosis. Modulators of CENP-E can thus control
 CC cell proliferation. Agents that modulate CENP-E activity are lead
 CC therapeutic, bioagricultural and diagnostic agents, e.g. for treatment

CC of unwanted cell proliferation (typical of many examples are tumors and
 CC metastases; vascular malfunction; inflammatory and immune diseases;
 CC angiogenesis; hypertension; restenosis; and fungal infections), also as
 CC plant-protection agents (selective herbicides, fungicides and
 CC insecticides) and plant growth regulators or activators for improving
 CC yields. CENP-E is also a diagnostic marker for dividing cells, including
 CC cancer cells.
 SQ Sequence 2954 AA;

Query Match 1.7%; Score 129.5; DB 1; Length 2954;
 Best Local Similarity 19.0%; Pred. No. 0.77;
 Matches 260; Conservative 190; Mismatches 441; Indels 479; Gaps 64;

QY 202 SDVGEFPVDRNAVQNNFMKDFSGKWTITSGLNPTRDAPDCQLHEFHMDNKLVLNLM 261
 Db 459 SDMPSPFEID--DSVCTEHSDFDALSMDSDNGIDAEN-----LASKVTH 502
 QY 262 RIKTIDGGFFTRSAVQTFVQDPLPGALVNHDEFLHYQDWYLLSSQIENKRPDYIFVY 321
 Db 503 REKT-----SLHOSMIDFGQISDSVQFHDSS-----KENQLOYLPRKD----- 540
 QY 322 YGRNDAMDYGGSVITYTRSPETIPESIIPLMOKAASVGRDNFTTNSGCPPEPLVE 381
 Db 540 -----SGDMAECRAKF--EKETISLQOQDQSKKEE-----KKEIVQ 574
 QY 382 RLE--KTAEERGEKLLIKEA-----VEIEVEKEVEKVRDTEMTLF----- 421
 Db 575 SFELKIALEBOLSVKAKNLEMTYNSRHSINAQVQDVEKEVYR--KEMSVLDGSGVN 631
 QY 421 -----ORL-----LEGFKELQO--DEENFYRELSE-----EKETI 448
 Db 632 ASNDLQDSSVDGKRLSSSHDECIEHRKMLQKIVLVEFEITENLKSSNDQKSSQEDOF 691
 QY 449 LNELOM--EATEVEKTLFGRALPIRKLRALAPHNSFLANETIKYYGSKLPGRKRFSGW 507
 Db 692 MESIQCEALMAEK-----ANALEELALMRDNPDIILNETIKRELAD----- 736
 QY 508 EDYGSIVYAKIGSSRIPIPRFKSRPICCGLDSSGLQLFSGKHNLSPANSI----- 561
 Db 736 -----LERSLKENOET--NEFELLEKETQHEHAQLIHEIGSLKLV 775
 QY 561 -----NQNPVPGNSGCKFPKFDVALMWEKQGFATAIVAFILISVASKDAVALATC 614
 Db 776 ENAMYNQNL-----EDLETK 792
 QY 615 TLLKEERLELACISNPACANVACLOTGNNRDETECQIR-----CGDL 660
 Db 793 TKLLKEDEIOLAE-----LRRRADNMQKVENFDSLVSMDSEKLCBEI 836
 QY 661 FE--NSVYDEFNCAVSRKKCYPRKSDVDFVPPPSVLYQKFMKDKDGGKWFITRGINP 718
 Db 837 FOLKQISLDAEATIRAOQKCSFLNSE-----NLELKEKEDRS--WYQNR--EK 883
 QY 719 TPDAPDCQLHEFTEENKLVGLNLSWRIRTPDGGFFTRSAVQFVQDPPYRGLVNHNEY 778
 Db 884 AASLFEKQLETERESNYKKMEADL-----QKELOS-----ARNEIN-Y 919
 QY 779 LLYQDDWYLLSSVENSPEPDYLFVYKGNNDAMDYGGSVLYTRSAVLPESIIPELOTA 838
 Db 920 LN-----GLIAGKV--PRDLL-----SRVLEKRVSEFSKOL 949
 QY 839 QKVRDFTFIKNDNTGCE-----PPIVERLEKLV-----EEGER--TIINEVE 881
 Db 950 EKALBEKNA--LNEVYTCSEYKFLPNEVBCLNQSKASEEMILKQGGESASISKOE 1008
 QY 882 ET--EEVEKVRKEVTLSEKLFEGKKELODEENFLRELSEENDVDGL--KMEA--- 935
 Db 1009 IIMQOSEQI-----IQLDEVYHTOSKYQTEEYGL--EMKRMHDLPEKYIRNKSSEADL 1063
 QY 935 -TEVEKLFGRALPIRKLMVATHCTSP-----CHRIFFSSDDGIGLGIITRKRIIN 986
 Db 1064 LREMENLKGTMSVEYKTIADTKHELETTIRKQELLHEKKYFOAMOTIFPI----- 1116

QY 987 GTEFLKLPISQADLRRTGGSSRP-----LSAFSGSGKI 1024
 Db 1116 -TFLSDSLPSKIVE-----GNSODPELEINDYNLALATERNNINWCCLETENSLKEV 1169
 QY 1025 FDI-VPLPS-----KNEIKELTAPLLKLV-----GVLAACFLIVPSA 1061
 Db 1170 IDINTQOSLOAQSIKESDLQKPRQDEGEVALLMELELLKHTLDSQSLTEKLOLENL 1229
 QY 1062 DAVDALKTCTCLLKGRICRIELAKCIANPACANACIOTCNRPDEPECQIKDCDLFEN-- 1120
 Db 1230 EYETKLOTLQDEKKNITIEEN-----LOT-NFEDLKAEHSLKQDLESEN 1275
 QY 1120 SYVDEFECASVRKRCVPRKSDGEPAPDPVLYONFNISDNGKWTYISGLNPTFADF 1179
 Db 1276 QSIETQDELAAQBELREQO-----LVDSFRQQLDCSVGIS--PHNDAY 1320
 QY 1180 DCQ---LHEFHTEGDNKLVG---NISWRIKLDSGFEFTRSAVOKFVQDPNQ--PGVLYN 1230
 Db 1321 ANQEKVSLGEVNSIQSEMLGERDELOTSCALVSELELLRAHVKSVEGENLEITRKING 1380
 QY 1231 HDNENLYHODWYIILSKINKPEDFYFYRRGRNAMDYGAVYTRSVLPNS--II 1288
 Db 1381 LEKRIKSESESVLKSMLENKED---NNKLKEQAEESKKNOPFLEEVSGSOKLV 1436
 QY 1289 PETE-----KAAS---IGRDFSTFIRIDNTGCPPEPALVERIEKIVEGERIIVKEVE 1338
 Db 1437 DEIVLAKQOLKAEEIRELEIDROVFELVQNTANNLVGKLETPLOADHEE-----DSID 1490
 QY 1339 EIEEVEKEV--EKVGRTMTLFOFLAEGFNEKODEENFRELKEME 1386
 Db 1491 RSEMEIKVLGEXLERNOY-LTERLOEKEKLELSNK-----LEILOKEME 1534

RESULT 13
 P50777
 ID P50777 standard; Protein; 1654 AA.
 AC P50777;
 DT 30-SEP-1991 (first entry)
 DE Sequence of the P195 protein of Plasmodium falciparum.
 KW Malaria vaccine; epitope; antigen; immunogen.
 OS Plasmodium falciparum.
 PN EP-154454-A.
 PD 11-SEP-1985.
 PE 21-FEB-1985; 301173.
 PR 22-FEB-1984; GB-004692.
 PR 26-SEP-1984; GB-024340.
 PR 21-FEB-1985; GB-004429.
 PA (WILL.) WELLCOME FOUNDATION LTD.
 PI Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
 DR WPI: 85-224845/37.
 DR N-PSDB: N50530.
 PT Cloned DNA sequence encoding plasmodium falciparum protein -
 PT useful for expressing the protein for use in vaccines against
 PT malaria
 PS Claim 6; Fig 1; 51pp; English.
 CC The sequence encoding the P195 protein of Plasmodium falciparum
 CC (N50530) and a peptide comprising at least one of its epitopes
 CC (see P50777) are claimed. Also claimed is a vaccine for inducing
 CC immunity to malaria comprising the novel peptide or P195 or a
 CC peptide comprising at least one epitope when derived from the new
 CC DNA sequence, together with a carrier.
 SO Sequence 1654 AA;

Query Match 1.7%, Score 129; DB 1; Length 1654;
 Best Local Similarity 17.3%, Pred. No. 0.34;
 Matches 238; Conservative 214; Mismatches 499; Indels 408; Gaps 67;

QY 184 VQGFECASVRKRCVPRKSDGEPAPDPVLYONFNISDNGKWTYISGLNPTFADFDC 243
 Db 249 IENINELLEESKRTDK-----NKNATKEBEKKKLVQAOVDS-----IYNK 290

QY 244 QLHEFHENDKLVGNLTIRIKTLDDGFEFTRSAVQTF--VQDP-----DIPGALYHNDE 295
 Db 291 QLEEAH-----NLISVLEKRIDTLTKNNIKELLDKIENIKPPANGNPNTLLDNKK 346
 QY 296 FLHYQDDWYIILSQIENKPPDY-----IFYVYGRNDAMDYGGSVYIYRSPFLPESII 349
 Db 347 IEHEKELEIKELAKTIKENIDSLFTDPLEEYLYLEKKNID--ISAKVETKESTEPNE-Y 403
 QY 350 PNLQKAASVSGDFENFTDSCGPEPPLVERLEKAEEGEKLLI-----KEAVEIEEE 404
 Db 404 PNGVYTPYST-NDINMLNELNSFG---DLINPDYIKESKKNITYTNERKKFLINELKEK 459
 QY 405 VEKEVEKVDFTMTLFOFLLEGKELODEENFVRELKKEKELINEL-----QNEAT 457
 Db 460 IKIEKKI-----ESDKSYEDRSKSLNDITKEKELINELIYSGFNNNIDLT 507
 QY 458 EYEKLFGR--ALPTKLMALAPNSNPLANETIKYVGSGLPCHKRPSGMWEDY-FGSI 514
 Db 508 NEKMMGRKRYSKVEKL-----THPNTFASYENSKHN--EKLTAALY--MEDYSLRNI 558
 QY 515 VYAKICSSRRITRYFRKSPRICGLDSKGLQFLSHGKNLSPASHINQNPKNSGC-- 572
 Db 559 VYKEL-----KYK-----NLKIKENI--ELVENIKKDEQLEF 595
 QY 572 -----KEPKDYALMWKMGQFAKTAIVALFILSVASKADAVALTCTCLKECHLE- 625
 Db 596 KITDENKPEKILEV-----SDIVQVQAVLLM-----NKIDELKKTQILNNVELKH 645
 QY 625 -----LAKISN-----PACAA-----NACIOTCNRPDET 651
 Db 646 NIHVENSTKOEKQEPYLYLAKKIDKLKYMFKVESLINEEKNKITQOSQNSPST 705
 QY 652 ECOI-----KCGDLFENSVDDEFECASVRKRCVPRKSDGEPVDPSPVLYO-- 700
 Db 706 EGEITGQATRKGGQAGSALG--DSVQAQAOEQKQNP-----PVPVVPPEAKQVP 756
 QY 700 -----KFDKNDTSKWFITRGINPTFADFQDLHEFHENKLVGNLSRIRTP 748
 Db 757 TEPAPVNNKTENVSKIDYLEKLY-QPLNTSYICHRYTLVSHSTNNELIKYK----- 809
 QY 749 DGGFFTRSAVQKTVQ-DPKYPIGLIYHNDENLYYQDDWYIILSKKENSPEDEYIFYVYGR 807
 Db 809 -----ITKEESKLSQDPL--DLFNIONTPV---MYSFDSINIYHNILWYIYE- 858
 QY 808 NDAMDYGGSVLY--YTRSAVL-----PESIIPELOTAQKVGDFNTFKTNTGPE 858
 Db 858 -----IGVYILMEIYERKMVNCILKMDONKIKKLBEAKKV-----STVKTLSSSMQ 909
 QY 859 P-PLVERLEKAYEBERT-----IKEYDEL-----EEVEKY 890
 Db 910 PLSLTPQDQPREVSANDDTSHSTNLNNSLKLFEINILSLGKNKIYQELIGQSSSENFYKI 969
 QY 891 RKEVYLFESKLFEGFELORDENFLBELSKEMOVLGLKEATEVEKLGFRALPIRL 950
 Db 970 LKDSDFYNESTNEVYKADINSINDSKK-----KLE-EDINKL-----KKT 1014
 QY 951 MAYATHCFTSPCHDRIRFSSDGGIGRGLGIRKRINGFLKILPPIQADLRITGGSS 1010
 Db 1015 LQSFPLYNKYKULERLDRKKKTVGKYMQIKTL--LTK-----EQLESXLNSLN 1064
 QY 1011 RP---LSARSQSKGIFDIPVLPKSNELKELTAPLLKLVYLACAFIIVSADAVAL 1067
 Db 1065 NPKHVLONSVFPNK-----KKEAE-----IATEENTL 1092
 QY 1068 KTCACLKGR--IELAKCIANPACANACIOTCNRPDETECOIKC--GDLFENYV 1122
 Db 1093 ENKTIILKHKG-VKYYNESSEPKTLSPESITQTDENYASLENFVLSLEKTLKDNINLT 1152
 QY 1123 DEFNECAVSR-----KCVPRKSDGEPAPDPVLYONFNISDNGKWTYISGLN 1173
 Db 1153 EKKKLSYLSRGLHLIAELKEVIAKKNYTGNSPS-----VANTDV-----N 1193
 QY 1174 PTFADFDCQLPFHTEGDNKLVGNLSMRIKTL--DSGFEFTRSAVQKFOVQDPNPQVLYNH 1231

Db 1194 NALSY-----KFLPECTD-----VATVSESGSDTLEQSO-----FKRPA----- 1231
QY 1232 DNEYLHODDWMYILSKIKENKPEDEYFYVYGRNDA--WDGAGAVYTRSSVY--PNSITPE 1290
Db 1231 -STHVAGASNTITTSQNVDEVDVIVLLFGESEEDYDLOGVYT--GEANTVSTV-- 1285
QY 1291 LEKAKSGRPFSTFIRINDNCGEPALVERIEKTVEEGRITVKEVEEIEEVEVEVER 1350
Db 1285 -----DN-----ILSKIE--NEYEVLYKPLAGYRSLSKKOLE- 1316
QY 1351 VGRTEMLFORLAEGFENELKODENFVRELSEKMEFLD 1389
Db 1316 --NNVMTFNVVVKDLMSRFKRNENFKNVLESLDLPYKD 1352

RESULT 14
R95268
ID R95268 standard; Protein; 993 AA.
AC R95268;
DE 01-AUG-1996 (first entry)
KW Nisin A; nisin gene; antimicrobial; preservative; antibiotic;
KW Lactococcus lactis strain NIZO R5.
OS Lactococcus lactis strain NIZO R5.
PN W09616180-A1.
PD 30-MAY-1996.
PF 20-NOV-1995; G02699.
PR 19-NOV-1994; GB-023404.
PA (BIOT-) BIOTECHNOLOGY & BIOLOGICAL SCI RES COUNC.
PI Dodd HM, Gasson MJ;
DR WPI; 96-268616/27.
DR N-PSDB; T29660.
PT Making cell which expresses nisin but does not contain natural nisa
PT gene - by providing cell with variant nisa gene, and genes for nisin
PT modification, secretion and immunity
PS Disclosure; Fig 7; 69pp; English.
CC The gene cluster nisA-nisP (see T29660 and T29651) of Lactococcus
CC lactis includes the nisa gene coding for pre-nisin A (R95267, see
CC also R95263) and the genes for nisin modification, secretion and
CC immunity. nisa (R95268) and nisc (R95270) are believed to be
CC involved in reactions that modify pre-nisin: nist (R95269) is
CC similar to a transport ATPase and is involved in translocation of
CC nisin out of the cell; nisti (R95271) is involved in immunity to
CC nisin. Replacement of the natural, chromosomal copy of the nisa
CC gene with a variant nisa gene allows produ. of high levels of nisin
CC A variants in Lactococcus lactis hosts.
SQ Sequence 993 AA;

Query Match 1.7%; Score 129; DB 1; Length 993;
Best Local Similarity 18.8%; Pred. No. 0.15;
Matches 125; Conservative 99; Mismatches 244; Indels 196; Gaps 28;

QY 297 LKHQDWMYILSSQIEKPPDY-----IFVYRGNDAMDGSGSVYTRSPITPESIT 349
Db 171 IKITNYOITSEFCENDYKEDICTVTLCYGDEYRESEQLYGLI-----VNHVLI 224
QY 350 PNLQAKASVGRDF--NFTITDNSCGPEPLV---ERLEKTAEGEKLKKEAVEIEEIE 404
Db 225 SNQKDLIS--DFSMTTLTKVEALDEDKYITPLKKYQKFIQYSEIEIGIEKLEKE 281
QY 405 VEKE-----VEKVRDTEMLT---FORLLEGEKE-----LOODEEN 436
Db 282 IQOEMSCILENDNYIQIDISSEINFVQKQOLEHLAEFLGNITTKSVRTYLDYDK 341
QY 437 FYRELSEKEKEIINELQMATVEVEKIFGALPIRKIRMLALHNSPLANHEITIKYVSK 496
Db 342 FLEKIVVDEVOITEL-----FDSTFGIGAPYNNH---PRNFISEPTLIVSEEE 391
QY 497 LQHKRFSWGEDYFGSIVVAKICSSRRIPRYFRKSPRICGLDSRGLOLFGHKNLSP 556
Db 392 R-----EKYLSMYEVA-----VKNHNININDD 413

QY 557 AHSINQVPGKNSGCKPFKQVALMWKMGOFATTAIVAFIL-----SVASKADAVD 609
Db 414 LESHQKM-----DLKSELOGLLEFLMLAKEYEKDFILSDIYGNNNLGGASGRFS 466
QY 610 AIKTCOLKRECELELAKCISNPACAN---VACLOTNNRPDETCQIKCGDLFENSV 665
Db 467 AL-----SPELTSYHRTIVDSVERENKREITSCELIV--LPEINR 505
QY 666 VDENECAVSRKCKVPRKSDVGDFPYDPVSLV-----QAFMDDESG-----KWFITRGL 716
Db 506 HANWHTSIMRKVLPFTSTSHNEVLITNLYIGIDEKEFYARDISTOELAKFYITSMY 565
QY 717 NPTDAPDCQ--LHEPHEENKLVGNLSWRIRTPDGCFFTR----- 756
Db 566 NKTLSFNSLRFLEYELSLDK--FENLPMELLYRFDYIPRLVPEIYISPAKKIMGRDV 623
QY 756 ---SAVQKQVDDPKYPGILY--NHDNEYLYQD---DWYILSSKVENSPEDYIFV---Y 803
Db 624 NSKMTIRRELQSKKEIPKEFYIVNGDNKRVYLSQENPLDMELLESARKSKRKDFIELQY 683
QY 804 YKGRNDAMDGSGSVLYTRSAVLPESTIPELOTFAQKVGHDNFTIKTDTCGPEPLVE 863
Db 684 FEDENIINKGKGRV-----ADVVPFTRRA--LGNEGRAFLIREKRV-----VE 727
QY 864 RLEK 867
Db 728 RREK 731

RESULT 15

ID R77087 standard; Protein; 1582 AA.
AC R77087;
DE 31-JAN-1996 (first entry)
DE Rat sulphonylurea receptor.
KW Persistent hyperinsulinaemic hypoglycaemia of infancy;
KW sulphonylurea receptor; diabetes; COS.
OS Ratus sp.
PN W09528411-A1.
PD 26-OCT-1995.
PF 12-APR-1995; U04463.
PF 13-APR-1994; US-226972.
PR 15-MAR-1995; US-404531.
PA (BAYU) BAYLOR COLLEGE MEDICINE.
PA (TBXA) UNIV TEXAS SYSTEM.
PI Bryan LA, Cole GJ, Gage RJ, Nelson DA, Thomas PM;
DR WPI; 95-373758/48.
DR N-PSDB; T01568.
PT Sulphonylurea receptors and corresp. nucleic acids - used in
PT detecting persistent hyperinsulinaemic hypoglycaemia of infancy
PS Disclosure; Page 99-104; 145pp; English.
CC Rat sulphonylurea receptor (SUR) DNA was ligated into eukaryotic
CC expression vectors containing the SV40 virus promoter and used
CC to transfect COS cells. Expression of the receptor was low
CC when SUR cDNA (T01551) was used, but high when the genomic SUR
CC sequence (T01568) was used; the product had the sequence given
CC in R77087.
SQ Sequence 1582 AA;

Query Match 1.7%; Score 124; DB 1; Length 1582;
Best Local Similarity 18.8%; Pred. No. 0.79;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

QY 868 KVEGERTIIEVEIEEVEKVRDEVTLLFSKLFEGFR--ELORDENFLRELSEKEMD 925
Db 480 KLSQAKRTIL-----EYSNRELKOT--NEMLAGIKLKLAMENIFCSRVEKTRXK 528
QY 926 VLQGLKMEA--TEVERLFGALPIRKILMAVAHCHTSPCHDRIKRFSSSD-----GIGRL 978
Db 529 EMTSLRAFAVYISISIFMTAIPAAVLITFV-----GHVSFFKESDFSPVAFASTL 580

```

QY 979 GTRKRINGTELL-----KILPPIO-----SADLRTTGGSSSRPLSAFRSGFSKG 1023
Db 581 SLFHLLVPLFLSSVSRSTYKALVSVOKLSEFLSSAFIREOCAPREPAPOGOA-----G 636
QY 1024 IFDIVPL-----PSKNELKELTAPLLKLVLGACFLVPSADAVDALKTCACLLNG 1076
Db 637 KYQAVPLKVVNRKRPAREVRDLGPLE-----RLTPESTDG-DADNFCVQIIGG 684
QY 1077 C-----RIELAKCIANPACANVACLOTCTNNRPDETECQKCGDLF 1117
Db 685 FFTWTPDGIPTLSNITIRIPRGOLTMIVGQVCGKSSLLATIG-----EMOKVSGAVF 738
QY 1118 ENSVVDENECVSRKCCVPRKSDLGEPAPDP-----SVLVQENFI-SDF 1162
Db 739 WNSLPD-----SEGRRPQOPRAGDSGRFGCOEQRPCGYASQKFWLLNATVEENITFESEPF 793
QY 1163 NGKMY-----ITSGLNPTEDAFDCQLHEFHTEGDNKLVGNISWRIKT----- 1205
Db 794 NKORXKWTIEACSLQPDIDL-----PHGDQTOIGERGINTLSTGGQRPDQCRPEPST 845
QY 1205 -----LDSGF-----FTRSAVOKFVQDPNPGVLYNHNDNEYLHYQDDWYILSS 1247
Db 846 STPMIVFLDDPFSALDVHLSLHMOAGILLELDRDKRTVVLVTHKLOYLPHA-DWIIAMK 904
QY 1248 KIENKPEDYIFVYRGRNDAMDGYGAVVYTRSSVLPNSIPELEKAASIGRDFSTEIR 1307
Db 905 DGTIOREGTLDFORESCQLEFHW-----KTIAMNRQDELEK-----ETVME 946
QY 1308 TDNTGPEPALVERIEKTEVEGERIIVKEVEEIEVEKE 1347
Db 947 RK---APEPS---QGLPRAMSSRDGLLDEDEEEEAASE 981

```

Search completed: October 14, 1999, 03:56:24
 Job time: 2707 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 1999, 03:19:10 ; Search time 33.23 seconds
(without alignments)
1702.452 Million cell updates/sec

Title: US-09-075-375-4

Perfect score: 7495
Sequence: 1 MALPASHNPLFANHEHTIKYVGSKLPGHKRFSWGMEDYFSGSIYVAKICSSRRIRPRFRKSP

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database: PIR_60:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2557	34.1	478	2	T03750 violaxanthin de-ep
2	2430	32.4	462	2	T00708 probable violaxant
3	135.5	2.1	886	2	H69378 purine NTPase homo
4	155.5	2.1	2829	2	A42771 reticulocyte-bind
5	155	2.1	2469	2	A36812 hypothetical prote
6	146	1.9	1935	1	A37102 myosin beta heavy
7	146	1.9	2748	2	S57976 nuclear migration
8	144.5	1.9	839	2	S54174 topoisomerase I -
9	144	1.9	2166	2	G70163 hypothetical prote
10	143.5	1.9	1558	2	B71603 RESA-H3 antigen PF
11	142.5	1.9	1935	1	S06006 myosin beta heavy
12	142.5	1.9	1819	2	A71928 cag island protein
13	141	1.9	1002	2	C70319 nitrate reductase
14	140.5	1.9	1302	2	UC6009 surface-located me
15	140	1.9	800	2	F64508 hypothetical prote
16	139.5	1.9	1025	2	S54044 probable membrane
17	137.5	1.8	1875	2	S38173 myosin-like protei
18	137	1.8	852	2	B43743 probable membrane
19	137	1.8	1676	2	E71410 probable centromer
20	137	1.8	1109	2	A40801 phosphoprotein pho
21	135	1.8	3660	1	S02041 dystrophin, muscle
22	134.5	1.8	1937	2	I38055 myosin heavy chain
23	134	1.8	1639	2	S05603 major merizolite su
24	134	1.8	978	2	A70387 conserved hypotet
25	133.5	1.8	1934	2	I48153 beta-myosin heavy
26	133	1.8	1979	2	C71622 hypothetical prote
27	132.5	1.8	1156	2	B70356 chromosome assembl
28	132	1.8	809	1	J00032 anthrax toxin leth
29	131	1.7	1156	2	E69444 chromosome segrega
30	130.5	1.7	1631	1	SAZOK1 major merizolite su
31	130	1.7	1679	2	S48385 hypothetical prote
32	129.5	1.7	1992	2	A47297 myosin heavy chain
33	129.5	1.7	1805	2	A64224 hypothetical prote
34	129	1.7	1201	2	A35815 myosin heavy chain
35	129	1.7	1175	2	C35815 myosin heavy chain
36	129	1.7	1822	2	S44849 K12H4.8 protein -
37	129	1.7	1179	2	F71190 probable chromosom
38	128	1.7	993	2	C31915 n18 protein - Lac
39	128	1.7	624	2	PC6003 surface membrane p

40	127.5	1.7	4910	2	S64942 probable membrane
41	127.5	1.7	1790	2	S67593 transport protein
42	126.5	1.7	2628	2	S59413 probable membrane
43	126.5	1.7	1282	2	JE0120 glycoprotein A - m
44	126	1.7	1640	2	A24594 probable major sur
45	126	1.7	993	2	C46951 n18 protein - Lac

ALIGNMENTS

RESULT 1
T03750
violaxanthin de-epoxidase precursor - common tobacco
C:Species: Nicotiana glauca (common tobacco)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
C:Accession: T03750
R:Bugos, R.C.; Hieber, A.D.; Yamamoto, H.Y.
J. Biol. Chem. 273, 15321-15324, 1998
A>Title: Xanthophyll cycle enzymes are members of the lipocalin family
A:Reference number: Z15054
A:Accession: T03750
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-478 <BUG>
A:Cross-references: EMBL:U34817; NID:q1463122; PID:q1463123
A:Experimental source: strain Xanthi; tissue-type leaf
C:Genetics:
A:Gene: TYDEL
C:Function:
A:Description: violaxanthin de-epoxidase and zeaxanthin epoxidase catalyze the additi
d in protecting the photosynthetic apparatus from excessive light
A>Note: established as member of the lipocalin family
F:1-134/Domains: transit peptide #status predicted <INP>
F:135-478/Product: violaxanthin de-epoxidase #status predicted <MAT>

Query Match 34.1%; Score 2557; DB 2; length 478;
Best local similarity 100.0%; Pred. No. 9.9e-151;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	474	MALPASHNPLFANHEHTIKYVGSKLPGHKRFSWGMEDYFSGSIYVAKICSSRRIRPRFRKSP	533
DB	1	MALPASHNPLFANHEHTIKYVGSKLPGHKRFSWGMEDYFSGSIYVAKICSSRRIRPRFRKSP	60
QY	534	RICCGDLSRGLJFESHGKHNLSPAHSINONVPRGNSGCKFPKDVALLMWEKKGQFAKTAI	593
DB	61	RICCGDLSRGLJFESHGKHNLSPAHSINONVPRGNSGCKFPKDVALLMWEKKGQFAKTAI	120
QY	594	VAFIILSVASKADAVALKTCCTLLKECRLELAKCISNPACAAVACLOTNNRPDETBC	653
DB	121	VAFIILSVASKADAVALKTCCTLLKECRLELAKCISNPACAAVACLOTNNRPDETBC	180
QY	654	QIKCGDLFENSVDENECAVSRKCCVPRKSDVDPPDPVSVLYOKFDMKDFSGKWT	713
DB	181	QIKCGDLFENSVDENECAVSRKCCVPRKSDVDPPDPVSVLYOKFDMKDFSGKWT	240
QY	714	RGNLPTFDATFVQLHHEHTEENKLVGNLSRITTPGGGFTTSAYOKFVODXPPIELYN	773
DB	241	RGNLPTFDATFVQLHHEHTEENKLVGNLSRITTPGGGFTTSAYOKFVODXPPIELYN	300
QY	774	HDNEVLLYDDWYILSKVNSPEDYIFVYKGRNDAMGYSVLYTSAYLPESITPE	833
DB	301	HDNEVLLYDDWYILSKVNSPEDYIFVYKGRNDAMGYSVLYTSAYLPESITPE	360
QY	834	LQTAQAVGSDFTFKTDNTGCPPEPLVERLEKKVEEERTTIKEVEEIEEVEKRD	893
DB	361	LQTAQAVGSDFTFKTDNTGCPPEPLVERLEKKVEEERTTIKEVEEIEEVEKRD	420
QY	894	EVLTFSLKEGFKELORDEENFLRELSTKEEMVDLDLKNHATEVEKLFERAPIRKL	950
DB	421	EVLTFSLKEGFKELORDEENFLRELSTKEEMVDLDLKNHATEVEKLFERAPIRKL	477

RESULT 2

T00708

Probable violaxanthin de-epoxidase precursor - Arabidopsis thaliana

N:Alternate names: protein F22013.3

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999

C:Accession: T00708

R:Shim, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Col
eologis, A.; Ecker, J.R.

A:Submitted to the EMBL Data Library, April 1998

A:Description: Genomic sequence for Arabidopsis thaliana BAC F22013.

A:Reference number: Z14200

A:Accession: T00708

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-462 <SHI>

A:Cross-references: EMBL:AC003981; NID:g3063438; PID:g3063441

C:Genetics:

A:Map position: 1

A:Mutons: 72/3; 128/2; 160/3; 292/2

Query Match 32.4%; Score 2430; DB 2; Length 462;

Best Local Similarity 100.0%; Pred. No. 6.5e-143; Indels 0; Gaps 0;

Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 MAAVATGCTSPCHDRIRFFSSDDGIGRLGTRKRINGTFLKILPPIQSDADLTGGRSS 1010

DB 1 MAAVATGCTSPCHDRIRFFSSDDGIGRLGTRKRINGTFLKILPPIQSDADLTGGRSS 60

QY 1011 RPLSAFSGSGSKGFDIVPLPSKNELKELTAPLLIKLVGLACAFILVPSADAVDLKTC 1070

DB 61 RPLSAFSGSGSKGFDIVPLPSKNELKELTAPLLIKLVGLACAFILVPSADAVDLKTC 120

QY 1071 ACLKGRGRIELACIANPACANVACLOTGNNRPDETECOIKGCDLFENSVDENECAY 1130

DB 121 ACLKGRGRIELACIANPACANVACLOTGNNRPDETECOIKGCDLFENSVDENECAY 180

QY 1131 SRKRCVPRKSDLGEPAPDPSVLYONFNISDFNGKWTITSGNLPTIDAFDCQHEHTHG 1190

DB 181 SRKRCVPRKSDLGEPAPDPSVLYONFNISDFNGKWTITSGNLPTIDAFDCQHEHTHG 240

QY 1191 DNKLVGNISMRITLDSGFTTSRVAQEFVODPNQPGVLNHDNEYLHYDDWTILSSKTE 1250

DB 241 DNKLVGNISMRITLDSGFTTSRVAQEFVODPNQPGVLNHDNEYLHYDDWTILSSKTE 300

QY 1251 NKPEDIYFVYRGRNDAMDYGAVVYTRSSVLPNSILPELEKRAKSGRDEFTIRTDN 1310

DB 301 NKPEDIYFVYRGRNDAMDYGAVVYTRSSVLPNSILPELEKRAKSGRDEFTIRTDN 360

QY 1311 TCGPEPALVERIKTYEEGRILLVKEVELEEEVEKEKVGTEMTLQRLAEGNEIK 1370

DB 361 TCGPEPALVERIKTYEEGRILLVKEVELEEEVEKEKVGTEMTLQRLAEGNEIK 420

QY 1371 QDEENFVRELSEKMEFLDEIKMEASEVEKLEFGKALPIKRV 1412

DB 421 QDEENFVRELSEKMEFLDEIKMEASEVEKLEFGKALPIKRV 462

RESULT 3

H69378

purine NTPase homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998

C:Accession: H69378

R:Flenk, H.P.; Claydon, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Flischiemann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Gloede, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997

A:Authors: Utebach, T.; Cotton, M.D.; Spiggs, T.; Ariach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Wose, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343

A:Accession: H69378

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-886 <RIE>

A:Cross-references: GB:AE001032; GB:AE000782; NID:g2689355; PID:g2649562; TIGR:AF1032

Query Match 2.1%; Score 155.5; DB 2; Length 886;

Best Local Similarity 20.6%; Pred. No. 0.063;

Matches 137; Conservative 85; Mismatches 221; Indels 221; Gaps 27;

QY 797 EDYFVYRGRNDAMDYGGSV--LYTRSAVLPESITPELOTAQKVGKGFNTFTD-N 853

DB 152 EDY-----ENAMKMLGAVIRMLEREKRLKEFLSQEHOIKRQEEK-----KAEIE 197

QY 854 TCGPEPALVERIKTYEEGRITL--IKVELEIEEVEKVRDKEVILFSLFEGFRELQ 910

DB 198 RISEBISIESLRKLESEVKNLESRLKELEHKSRLSELSRKOESSVLOEV-----R 249

QY 911 DEENFRELSEKENDV--LDGLMEATEVEKLFGRALPIRKIMAAVATGCTSPCHDRIR 967

DB 250 GLEEKRLRELEKQLEVERIEDLEKKAKEVKEL-----KPKAERY- 290

QY 968 FFSSDDGIGRLGTRKRINGTFLKILPPIQSA--DIRTTGSSRPLSAFSGSGSKIT 1025

DB 290 -----SILEKLSEINOALRDVEKREGDLTREAGIQOLKAAE 329

QY 1026 DIVPLPSKNELKELTAPL-----LKLTVGLACAFILVPSADAVDLKTC 1070

DB 330 D-----NSKLEITTKRIEELERLEFEKSHRLE-----TLKPRMDNMOGIK-- 373

QY 1071 ACLKGRGRIELACIANPACANVACLOTGNNRPDETECOIKGCDLFENSVDENECAY 1130

DB 373 -----AKLEKNLTPDKVE--KMYDLSKAKEEE-KEITE 404

QY 1131 SRKRCVPRKSDLGEPAPDPSVLYONFNISDFNGKWTITSGNLPTIDAFDCQHEHTHG 1190

DB 405 KLKRLINKSSLSKTRGQOLKAAEE-----LKSARTCPVCGRELDDEHRRN 451

QY 1191 -----DNKLVGNISMRITLDSGFTTSRVAQEFVODPNQPGVLYN-- 1231

DB 452 IMAEYTRMKRIAELEKADIEKKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLE 511

QY 1231 --HNEFLHODDMYITLSSLEKPEDYFVYRGRNDAMDYGAVVYTRSSVLPNSIT 1288

DB 512 SSHDAE-----KLSAESE--YRKVERLDGLRGQOKILLSSA--SRI 550

QY 1289 PELEKAKSGRDEFTIRTDNCGPEPALVERIKTYEEGRILLVKEVELEEEVEKEK 1348

DB 551 KELKSSUREI-----EAL-KVYESRGELHKRIEEDGESLEELEREV 593

QY 1349 EKVGRTEMTLQRLAEGFN--ELKODEENFVRELSEKMEFLDEIKMEASEVEKLEFGRA 1405

DB 594 -----QSLRPYNNKMLKLEKDAESRLSEELKRR- KLEDELSBALAKLEANGRA 641

QY 1406 LPIR 1409

DB 642 EDIR 645

RESULT 4

A42771

reticulocyte-binding protein 1 - Plasmodium vivax

C:Species: Plasmodium vivax

C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995

C:Accession: A42771

R:Ghalinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992

A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.

A:Reference number: A42771; MUID:92315358

A:Accession: A42771

A>Status: preliminary.

QY 1115 DLFENSVDDEFNECAVSRKKCVPRK-SDLGEFPADPDSVLVONENISDPNGKWTITSLN 1173
 DB 1803 HL--NLVLLDQNESTYLRPLENREVEDLINTLSKLNITAIPTTYODIGKYE-----N 1855
 QY 1174 PTFDADFCOLHFHFHTEGDKRLVGNISWRIKLDSGFFTRSAVQKFEVDPNPGVLY----- 1230
 DB 1856 PNFDIYAKDSLKN-----MDYVAISNODYELMVAAYEKQOLDYLTIS 1896
 QY 1230 -----NHDNEYLHYODDWYIISSKLENKPEDIYIYYIGRNDADG 1270
 DB 1897 SERKIDIVVPLSEYNLMVNTYRNPSSLYLKEKAVLNHILIKEDDYKNTL-----AVSE 1950
 QY 1271 YGGAVVYTRSSVLPNSIPELEKKAASIGRDPSTFIRDTNTGPPALVERIEKTEVEEGE 1330
 DB 1951 HPTVHLSERKASLKNVLDKD-----DRAITMSRSLIE--KPTIDELSTKALSMG- 1998
 QY 1331 RIIVKEV-----EEIEEVEKEVEKVRTEMTLFORLABGFNELLKODEENFVRELKSEEM 1385
 DB 1998 KILVNSTHKNRKNKLSSEDPSEFLTMKAKEOGLIISKEYSELROQIDRPSLDYLKEKA 2057
 QY 1386 EFIDELKMEASVEKLFPGKALP 1407
 DB 2058 AIFDSIIVENIEYQOLVNTTSP 2079
 RESULT 8
 S54174
 topoisomerase I - Plasmodium falciparum
 C:Species: Plasmodium falciparum
 C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
 C:Accession: S54174
 R:Tosh, K.T.; Kilbey, B.J.K.
 Submitted to the EMBL Data Library, January 1995
 A:Description: Isolation and characterisation of the topoisomerase 1 gene from Plasmodium
 A:Reference number: S54174
 A:Accession: S54174
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1839 <TOS>
 A:Cross-references: EMBL:X83758; NID:g790481; PID:g790482

Query Match 1.9%; Score 144.5; DB 2; Length 839;
 Best Local Similarity 17.3%; Pred. No. 0.28;
 Matches 191; Conservative 176; Mismatches 390; Indels 347; Gaps 46;

QY 352 LQKAASVGDHFNNTTDSGPEPLVERLEKTAEGEKLKAEVET--EEVEKEV 409
 DB 24 INKIKQMLG--NN--KSCNSRSSKSKESIKKOKNSNSELGIKKNTKKSLGIKKEEEKKQI 78
 QY 410 EKVYDTEMTLFORLEGFKELODEENFVRELKSEKEITNELQMEATEVEKLFGRALPI 469
 DB 79 SKRSNLEKERNKLEKGGKKVVEKSKRTV-----KDETKLINVIKKEVQNNKK-----PK 128
 QY 470 RKLMLALAPSNFPLANHEIKYVSKLPGHKRPSWGMEDYFGSIVYAKICSSRRIRPYF 529
 DB 129 KLLAKS--EENF--EPINRW--WEKIDQDTDIQWNY----- 159
 QY 530 KRSRPRICGDLDSRGLQLEFSHGKHLMSAHSINOVVPGNSCKRPPKVOALMWEKMGQFA 589
 DB 159 -----LEHRLG-IF-----SPRY-VQHHVP-----IFYKIKIEL-----NA 188
 QY 590 KTAIVAFILSVASKADAVDALKTCTCLLKRCRLELAKCISNPACANVACLOTQNNRPD 649
 DB 189 KSEELATWCS-----AIGSDYCKEKEFILNFFETFIN-----SLENDNITIKQ 231
 QY 650 ETECOIKGDLFENSVDDEFNECAVSRKKVPRKSDVGDFVPDPSVLVQKFDN----- 704
 DB 232 ENETKLLKGGDISNFKFID-----FMPKIDHLLKLEKINKTKREE 271
 QY 704 KDFSGKMTITGLNPTFPAFDCQJLHFFTEENKLVGNISWIRINRPPDG-----FFTSAY 758
 DB 272 KEERKMRERKELPYTVALVDMIREKISSKNAEPPGLFRGGEHPKQGLLKKRIFFEDVY 331

QY 759 QKFVODPKYDGIILYN-----HDNE--YLLYODDWYIISSKVENSPEDIYIYYIKG 806
 DB 332 INSKDAPVPRNLDNMGHMGDIYHDKNTWYLAIVKID-----SINQIKTFLS 381
 QY 807 RNDAMDYGGSVLYTRSAVLPESIIPELOTAOKVRDENTFIKT---DNTCGPEPLV 862
 DB 382 AQSFKGYKDKLMKYENAR-----KLKSCVHKIRIEDYKKNKKNKNIIDKOLGTAVYLI 433
 QY 863 ERLEKKVEEERTLIKVEEIEFEVEKVRKVTLSKLEFGKELODEENFVRELKSE 922
 DB 434 DFLALRV-GGER-----DIDEADIV-----GCCSLRVEHISFAHDIPK 472
 QY 923 EMDVLIDLKMEATEVEKLFGRALPIRKLMVAARHCTS---PCHDIRFFSSDGIIGRLG 979
 DB 473 SYD-----SKDKTDEKVNKIPLPTNESISSDQYITLDFLGKDSIRYENT----- 521
 QY 980 ITRKRINGTEFLKILPIPIQSADLRTTGGRSSRPLSAFRSGFSKGIIDYVLPBKNL-KE 1038
 DB 521 --VKIDKQAVINII-----IFCKNNRDEGVDPDQITCSKLNELYKE 559
 QY 1039 LTAPELLKINGVLACAPLIVPASADAVDAL--KTCACLLKGCRIELEKCIANPACANVAC 1096
 DB 560 IMPTLSAKVEFTIVASITLDOQLKRIEYVGTITSLYSG-ETELRK-----SK 607
 QY 1097 LOTCNNRPDETECOIKGDLFENSVDDEFNECAVSRK---KCVPRKSDIGFEPADPSV 1152
 DB 608 KRKSHLSDINILSDASDSTINDVNNEDENGINKLSTVATYTGKENDVDKNSP-IEV 666
 QY 1153 LVQNFENISDENGKAVITISGLNPTFADPCQJLHFFTEGDKNKLGNISWRIKLIDSGETR 1212
 DB 667 DVSNIN-----ELINYNANNEVALICHOHSIPKQHDITM 703
 QY 1213 SAVOKFVODPWGfGVLYNHD-NEYLHYODDWYIISSKLENKPEDIYIYYIGRNDADWG 1271
 DB 704 SKIKQIE-----LYNEDIKEIKKY-----LOHLKNSDKKFIIV----- 729
 QY 1272 GGAVVYTRSSVLPNSIPELEKKAASIGRDPSTFIRDTNTGPPALVERIEKTEVEGER 1331
 DB 739 -----SKVSTLDGTLR-----NKKVENMK 758
 QY 1332 ITVKEVEIEFEVEKEVKGVRTEMTLFORLABGFNELLKODEENFVRELKSEEFIDE 1391
 DB 759 -----ESCKRKL-----ITLIKVELLNQMKVADKNTALGTSKINIMDPR 802
 QY 1391 -----IKMEASEVEKLFPGKALPIR 1409
 DB 803 ITVACCKKFEITIEKVFNRSLRK 826

RESULT 9
 G70163
 hypothetical protein BB0512 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
 C:Accession: G70163
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Tachigra, R.; Wh
 son, D.; Peterson, J.; Kerlavang, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: G70163
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2166 <XEF>
 A:Cross-references: GB:AE001153; GB:AE000783; NID:g2688419; PID:g2688426; TIGR:BB0512
 A:Experimental source: strain 831

Query Match 1.9%; Score 144; DB 2; Length 2166;
 Best Local Similarity 16.8%; Pred. No. 1.2;

Matches 268: Conservative 275: Mismatches 556: Indels 500: Gaps 71:

```

QY 44 RSNNGYNSRFLPTSYTSSFS--DSSHCKDKQICSIDTSEFIOQFLDKRM---TLI 98
DB 482 RNNYNSLNDINIAKYALFESIDSSSSKFENOMESKYSFTD---KTAAMDSESLM 537
QY 99 LKQWQFIOALVCTGVIVPRVAVDALKTCACCLKCRLELAKCIANPSCAANVAC 158
DB 538 YGEFFELSO-----EATNVOEPQDINKRLNE-----567
QY 159 LQTCNNRPDETEQIKCGDLFENSVDQFNECAVSRRKCYPRKSDYGEFFPRDANAVON 218
DB 567 IESYNNFEKTOETLK-----VD-FNTSLIN-----IDELCK-----599
QY 219 FNNKDEGKAVITSGNPTFADQCQLEHFMENDKLVNLTWRKTLDSGFEFTSAVOT 278
DB 599 -NVEFRDRY-----DEVNIFVTOLEESKIQYSKWQGMOSNLKNIE-----641
QY 279 FYODDPLGALYHNDNEFLH---YODWYILSSQIENKPDYIFYYRGRANDMOCYGG 334
DB 641 -----SQINKTNEFFLSLQIDKDGIELSESVFNDSHT-----OKKAIMHGS 686
QY 335 ---SVITRSPYL-----PSSILPNLOKAAKSVGRDNFTTDSNGEPPPLVERLE-- 385
DB 687 WKDELALNKLSDIKVSSSEELSSATLKIESLEKDVN-----DMEYV 730
QY 385 -KTAEEGKLIKKEAVE-----IEEEVEKEKAVDTMTLFOQLBSGF 427
DB 731 LKKTGDI-ESVLEYEKELKDKMSYSDSAIIGIKFINKRQTEILLKDKSVFMLEDNKKEF 789
QY 428 KELQDQENFV-----RELSEKEKELNLOMEATE-VEKLFGRALPIKRLAMA 475
DB 790 ---DDKNNEFVISKIECDYKDKFIESEDILNFKSPLNEFIES-----KLOIV 836
QY 476 LAPH------NFLANHETIKYVGSKLPHKRRPSWGMEDYFGSIYV--AK 518
DB 837 SNKSNOKQIDDELDRISKDILNRKDSINNEVDSKLS-----DMOSKLEITVKIEN 889
QY 519 ICSSRRIP-----RYFRKSPRICGSDSRGLOLFHSGK-----552
DB 890 LSSGKVDLDLIDSEVTTIKELKFSIES-----LESYLLKIDEFKOGAIVSDELLQ 943
QY 552 ---HNLSPAHSHINQVPGK---NSGCKFEPKVALMWEKQOPAK-TATVAIFILSV 601
DB 944 DIMNHENKTRLEENLSKFAAVLNSEEFKVEVDSLLQDKRTDIASQANIDITLDSL 1003
QY 602 ASKADAVDALKTCTCLKCRLELAKCIJN-PACAAVAVC-----641
DB 1004 NVRKFNIN-----KINQKYNVEVJISNYGISENISSKLENIHMEIENLSRRLDR 1054
QY 641 -----LQTCNNRPDETEQI---KCGDLFE-----NSVDFENECASVR-K 677
DB 1055 IDLSKGMENLOKLESFDSVYKQVEKFLKVDLTDOGEAKIKLVLKEIQYKKSLE 1114
QY 678 KCVPRKSDYGEFPVDPVSVLYVQKFMKDFSGKWFITRGLNPTFADFQCLHEPHTENKL 737
DB 1115 EADIRRTI-----DNDIM-----QAKERFGE--ITNELKNNIESKSEFLNDLYKERFKL 1162
QY 738 V-GNLSMRIRTPDGGFRTSR-AVQKFWODPKYPGIIYNHNEVLYLQDDWYIILSKVEN 794
DB 1163 IESNEEFNST---FLIESEGAISK-IRDELYKTLTSDEN-----LOIKISE 1206
QY 795 SPEDYIFYYYKRN-----DAMDYGGSVLYTRSAYLAPESIIPELOTAQKV 841
DB 1207 MDQNEEIIEORSKDLFEKELODKIKDCY-GFINSQGEIKAGVEENIKNHPVCJKV 1265
QY 842 GRDFTFTKTNDTCGPEPPLVERLE--KKVEGERTITIKV-EELIEVEVAVRD---KE 894
DB 1266 ---NTLI-DDDIYAVENELIKRIDSLSKISSTPDSIEKNLNDVSGCIDKIANDEMLKY 1320
QY 895 VTLSEKLEGEKELQDEENFLRELSEKEMDVLDGLKMEATEVEKLFGRALPIKRLAMA 954
DB 1321 IELERCNEGQNLLENKIDNKIKAMIDNALSQYDGLKRYADMDFESERL--NSYIATL 1378

```

```

QY 965 THCFSPCHDRIRPFSSDDIGRLGITRRKNG-----TFULKILPIQ 998
DB 1379 SEEFSSNKKEMI--FELESQKLNKLNESDNNVNEKVINLKKEESHVNSHLKEEDF 1436
QY 999 SADLRTTGGSSRRLSARFSGFSKGFIDIVLPSPKNELKELTAPLLKLVGLACAFIIV 1058
DB 1437 FKDLIRGEELKYSLENFNLSYNDKIQNLFEYDLSKN-----1473
QY 1059 PSADAVDALKTCACCLGCRLELAKCIANPACAAVAVCLOTCNNRPDETEQIKCGDLFE 1118
DB 1473 -----LENKTLFLOSFRDLIEQKKRDKENFYDFTKEFSKKKDMOSEIA---LME 1521
QY 1119 NSV---VDEFNCAVSRRKCVPR-----KSDLGEFPAPDSVYVONFNISDF-----NG 1164
DB 1522 TNITGKDEFDYFNKNGOSIIDSFWFLNKIDVADMOKESITIEKRINLAELGIRKFEND 1581
QY 1165 KWTITSGINPTFADFQCLHEPHTEGDN-----KLVGNIISWRKTLDSGFTT 1211
DB 1582 IFNVKIGLESEKDFELIKABEIFSNLQNEAKKIEOSVHLDKFNIGE-SINKLVLDLEKTV 1640
QY 1212 RSAVQKFWODPNQ--PGVLYNHNDXYLHYOD-----WYLLSKIEENKE 1254
DB 1641 DFKLEKIDEXYNNKTEIDILLQAEVKFLTOOKDLEDKIFELNOKLEHEFTLSSNDKVR 1700
QY 1255 DYIFVYRGRNDAMDYGAVVYTRSSV-----LPNSIIPLEKRAKSGIRD 1301
DB 1701 EMDVVISDR---ESPEGQIELNKNISFSEKISLYRNNIETSIENYNSPSKSKISD 1756
QY 1302 FSPFTFNDNCCPEPALVENIEKTEVEGERIYKVEEELIEBEVEKVEKVGRTMTLFO- 1361
DB 1757 LG-----LEDELKLSLNHS-----TSEIETTKSGLOEIDK-----FEV 1791
QY 1361 RLAEGFNELODEENFYRELSEKEM-----EFLDEIK 1392
DB 1792 EFKKNHKEILKEVDNNILEESKILNCDVOFNKFISEIK 1830

```

RESULT 10

B71603
RESA-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 13-Nov-1998
C:Accession: B71603
R:Gardner, M.J.; Teitelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H. Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600
A:Accession: B71603
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Restrictions: 1-1558 <GAR>
A:Cross-References: GB:AE001424; GB:AE001362; NID:q3845307; PID:q3845309; TIGR:PFB091
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0915w

Query Match 1.9%; Score 143.5; DB 2; Length 1558;
Best local Similarity 17.9%; Pred. No. 0.77;
Matches 206; Conservative 153; Mismatches 377; Indels 413; Gaps 50;

```

QY 347 STTPNOK-AASVGRDN-----NITTDNSCGPEPPLVE-----RLEKTAEG 390
DB 626 TILNEVEETASVTFNSILIEIQENTITNDI---EKLLEELHENYLSALENTQSEE 682
QY 391 EKLLEAVEIEEVEKEV-----EKVADTEKTLFQRLLEGKELQOD---ENNVAREL 441
DB 683 EK---KEVIDVIEVKEEVAITLITVQDAEESASTITTEFENIEEAVNSNEVAVENL 739
QY 442 SKEEELNLOMEATEVEKLFGRALPIKRLMALAPHSNFLANHETIKYVGSKLDPGRK 501

```

```

Db 740 EKLNETVFNVLKVEETVEISGSLNNEMDKAF-----FSELPDNVKGIOENLTLG-- 793
Qy 502 RFSMGWEDYFGSIYVARIKSSRRIPRRKSPRICGLDSRGLOLSHG---KHNLSPA 557
Db 793 -----MFRSIFTSIYIOSEE--KYDLNENYVSSITLDN--IENMEKGLNKLLENISST 840
Qy 558 HSIQNPYKPG--NSGCKFPKVDALMWEKMGQFATAIYALFILLSVASKADAVALKCTC 616
Db 841 EGVQETVTEHVEQNVYVDVDPARK-----DQF-----LQILNAGG----- 878
Qy 617 LKRCRLLELACISNPACAAVACLQTNRPDETECO---IKCGDLFENSVDNECA 673
Db 878 -LKEMFNLLEVEFKS---ESDVIIVEELIDEPVQKEVEKEVSIIEEMEEIVDVEE-- 932
Qy 674 VSRKKVPRKSDVD--FPVPDPSVLVQKFDKMFSGKWFITRGLNTPFDAQCQLHEFH 731
Db 932 -----EKEDLTKMDIDAVEESIEISS--DSKEET-----ESI 961
Qy 732 TEENKLVGNLSMRIRTPDGGFTTSAAQKFPVQDPKYGILYHNDNELLYQDDWYILSK 791
Db 962 KQKEDVSLVVEEVDND---MDESVEKYLE-----LKNMEELMKDAVEIINDITSK 1010
Qy 792 -VENSPEDIYFYVYKGRNDAMDGYGSLYTRSAVLPEST--IPELOTAQKVRDENTF 848
Db 1011 LIEETQE-----LNEVADLTKMDKELKELEKA--LSSEDSKEI 1046
Qy 849 IKTDNTCCPEPPLVERLEKKEVEGE--RTIIEVEET---EEVEKYRD-----K 893
Db 1047 IDAKD-----DLEKVEIEEHDITITLDEVELKDVEDKIEKVSJDKLEEDITLK 1097
Qy 894 EV-----TLESKLBEFGKEJONDEENFRELSEKEMDVLDGKMEATEVEKLEFGRAL-- 946
Db 1098 EYKEIKELSELLEPYKELKTITDLIEKKEIKEDHKEKEEKEEELKDLADILKVS 1157
Qy 946 -----PIRKLMAVAHCFSTSPCHDRITRFSSDDGIGRLGITRKRRINGFLFKIL 994
Db 1158 SLEVEEKKLEBEVHLKEVEHHIISGDH-----IKG----- 1190
Qy 995 PRIOGADLRTTGRSSRPLASRGSFGIPIVLPSPKNEKLELTAPLLKLKLVGLACA 1054
Db 1190 --LEEDDEEVDLKGSIIDMLKGMELGDMD-----KESLEDTAL-- 1231
Qy 1055 FLIVPSADAVDALKTACILKGCRIELAKCIANPACAAVACLQTN--NRDETECOIK 1112
Db 1231 -----GERVESLKDVLSALSALGMEDEGMKTRKAQRKLEBEVLK 1269
Qy 1113 CGDLFENSVDNEFNCAVSRRKCVPRKSDLGEPAPDPSVLVQNNISDFNGKWIITSG 1172
Db 1270 -----EEVKEEPPKKRITKKK-----VRDIDK-----K 1292
Qy 1173 NPTFPAFDCQLHEFHTEGNNKLVGNTISWRIKTLDSGFTFRAVOKFOPNPGVLYNHD 1232
Db 1293 EPEDEIVEVEKMDDEDID-----IEEDVEED----- 1320
Qy 1233 NEXLHAYQDDWYILSKRIENKPEDIYFYVYGRNDAMDYGAVYTRSSVLPNSIPELE 1292
Db 1320 -----IEEDKVEDIDEDID-----E 1334
Qy 1293 KAAKSGR-----FSTFIRTDNTCCPEPALVERIEKTEVEGERILYKVEE-----IE 1341
Db 1335 DIDEDIGERKDEVILDIYQKEKRIEYKKEKKLEKKEVSGKLEKHDEVMKYVOKID 1394
Qy 1342 EYVEKEVEKY--GRTMETTLFORLAEFNLKODE-----NFVRLSKSEEMFLDEIM 1393
Db 1395 KEVDKEVSKLSEKNDVT-----NYLKQNDFFSKVKNVKKYKAYFAAFISAVAA 1445
Qy 1394 EASEVEKLF 1402
Db 1446 EASYVVGFF 1454

RESULT 11
S06006

```

```

myosin beta heavy chain, cardiac muscle - rat
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence,revision 31-Dec-1993 #text,change 13-Feb-1998
C:Accession: S06006; S07536; 167441
R:Kratz, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
Nucleic Acids Res. 17, 7529-7530, 1989
A:Title: Complete nucleotide sequence of full length cDNA for rat beta cardiac myosin
A:Reference number: S06006; MUID:90016823
A:Accession: S06006
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-1935 <KRA>
A:Cross-references: EMBL:X15939; NID:956656; PID:956657
R:McNally, E.M.; Kraft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
J. Mol. Biol. 210, 665-671, 1989
A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Compari
A:Reference number: S07535; MUID:90133919
A:Accession: S07536
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-950, 'RK', 953-1935 <MCN>
R:Mandau, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.
Eur. Heart J. 5, 181-191, 1984
A:Title: Cardiac myosin heavy chain isozymic transitions during development and under
A:Reference number: I1305; MUID:85179510
A:Accession: I67441
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1871-1935 <RES>
A:Cross-references: GB:M32698; NID:9205598; PID:9205599
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methyl
F:88-766/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:348-585/Region: actin binding #status predicted
F:655-677/Region: actin binding #status predicted
F:839-1935/Domain: coiled coil #status predicted <COI>
F:839-1279/Region: S2
F:1280-1935/Region: light meromyosin
F:129/Modified site: N6,N6,N6-trimethyllysine (lys) #status predicted
F:184/Binding site: ATP (lys) #status predicted
F:695/705/Active site: Cys #status predicted

Query Match 1.9%; Score 142.5; DB 1; Length 1935;
Best Local Similarity 16.7%; Pred. No. 1.2;
Matches 247; Conservative 218; Mismatches 558; Indels 459; Gaps 56;

```

```

Qy 166 PDTECOIKCGTIFENSVDNEFNCAVSRRKCVPRKSDNGEFPVPRN-----AY 216
Db 375 PDGTE-----FADKRSAYLGMLSADLLKGLCHPRKVGNEYVTQKQVAYAGALA 428
Qy 217 ONFMKDSFGKTYITSLNPT-----FDADCOLHE-----FHMENDKL 255
Db 429 KSVYERKEN--WNYTR--INATLETQPRQYFTGVLDIAGEIFIDFNSFDCLINTNEKI 485
Qy 256 VGNLTWRITKTLGGFTFSAVQ--TFVQ-----D 282
Db 486 QQFNNHMFVLEDEYKKEGIEWTFIDGMDLQACIDLIEKPMGIMSLIEECMPKATD 545
Qy 283 PDLPGALY--NHNDFFLHAYQDDWYILSSQIENKPEDIYFYVYGRND-----AMGYGSGVI 337
Db 546 MTFRAKLYDNMLIGKSNFQPRNIRKQEAH-----FSLHYAGTVDNILGW-----L 594
Qy 338 YTRSPFLPESTIFNLQKAA--KSVGRDNFNTTTDNSCGPEPPLVERLEKTAEGEKILLIK 396
Db 595 QNKDPLNFIYVGLYQKSSKLISNLNANY-----AGADAP--VDKKGAKKAGSSSGTV 647
Qy 397 EAVLEEEVEKVEKVRDTEMTLIFORLLEGFKLQODEENFVRLSKKEE----- 448
Db 648 SALH-REVLNKLMTNLRST-----HPHFVRCIIPNETKSPGVMDNPL 688

```

```

QY 448 IINLOMEAT-EVKLFGRALP-----IRKLMALAPHSNPLNHEHTIKYVGS 495
DB 689 VHHQRLCNCVLEIGITICRQGFPPNRLLYDGRQRYRLPLNPAIPEGOIFDSRKGAEKLGS 748
QY 496 KLPGRKRSWGWEDYFAGIVAVAKICSSRRIPRYFRKSPRICGGLDSRGLOLFSHGKHLNS 555
DB 749 LDIDHNOYKFGHTKVFFAGLGLLEEMRDELSRIIRI--QASRQVLSRMEFKLLE 806
QY 556 PAHS---INONYPKGNSSGCKFPKDVALLWWEKGOFAATAIYAFISVASKADAV---- 609
DB 807 RRDLSLLIOWNIR-----AFMGVKNWPMKLYFKIKPLKSAFTEKEMANKE 854
QY 609 -----DALKTCTCLKCRLELAKICSPACANVAVLOJCNRPDETCQICGLDFE 662
DB 855 EFGKVKDLSEKSEARKELEKRWMSLDFE---KNDLOLOVABODNLDAEERCDOLIK 910
QY 663 NSV-----VDEFNE-----CAVSRR---KCVPRKSDVDFPVPDPSVLQKFD 702
DB 911 NKIOLEAKVKEMTERLEDEEEMNALLTAKRRKLEDEGSELKRDIDLELTLAKYERK-- 969
QY 703 MKDSGKMFITRGINPTDADFQOLHEHTENKLVGNLSKRITPPGCGFTSRSAVOKEV 762
DB 969 -----HATENK-VKNLLEEMAGLD-----EIIYKLT 993
QY 763 QDPKPGILYNHNEYLLODDWYLLSKSVNSPEDEYFYVYKGRNDAMPDGYGSVLYTR 822
DB 994 KEKKALEAHOQALDDQAEEDKVNLTAKAVKLEQV-----DDLESGLDOK 1042
QY 823 SA-----VLPESITPELOTAQKVG---RDF-----NFIKTNTGCP 857
DB 1043 KYRNDLERAKKLEBDLKLQESITMDLNDKQOLDERIKKKDFELNLRNIEDEQALGS 1102
QY 858 E-----PLVERLEKVEEGERT-----IKVEEIEEVEK----- 890
DB 1103 QLOKKLELQARIELEBEL- EAEIRAKAVEKIRSDLSLELEISRLLEAGAGISVOI 1161
QY 890 -----VRDKEVTLFESKLEFEGEKELORDENEFLREL 919
DB 1162 EMNKKRAEFOKMRDLEEAATLQHEATAALRKHADSVAFELQJDNQVAKLEK 1221
QY 920 SKEDAVLDGLKMEATEVEKLFGRALPIRKLMNAVATHCIFSPCHDRIRFSSDQIGRLG 979
DB 1222 SEPLELDD---VTSMEOTIRAKANLEK---CRF--LEDNNEHRS-----KAB 1264
QY 980 ITRKRINGITLKLIPRIGSADLRTTGGSSRPD---SAFRSGSGIGDIYLPKSNEL 1036
DB 1265 ELOGSVNDL---TRAKALQTEGELSRLQDEKALISOLTRG---KLTYYOOL 1312
QY 1037 KELTAPLKLTVGLACAFILVPSADAVDLKTCACILKGCRIELAKCIANPACANVAC 1096
DB 1313 EDLKRQLEEEVAKKANALALQASARHCDLLRQYEEETEAKAEIQRVLSK--ANSEVAQ 1370
QY 1097 LQTCNNRPDETCQIKGDFJFENSVDDEFNECAVSRRKCVPRKSDGGEPPADPSVLYGN 1156
DB 1371 WRT-----KYETDAIQRTELELEAKKKAQRIQDDEF----- 1403
QY 1157 FNTSDNGKMYITSGNPTDADFQOLHEHTENKLVGNISWRIKTLDSGFTSRSAVO 1216
DB 1403 -----AVEAVANAKSSLEKTKHRLONEIEDLMDV-----VERSNAA 1438
QY 1217 KEVODENQGVILYNHNEYLHODDWYLLSKIKENKPEYIIFYVYGRNDAMPDGYGAVY 1276
DB 1439 AALADKKQ---RNFDKILVEKKQKYEESOSELESSQKE-----ARSL 1477
QY 1277 YTRSSVLPNSITPELE-----KAASIGRDFSTFRTDNTGPEPALVERTEKVEE 1330
DB 1478 STELEFKLNAVESEHLETFKRENKNLOEISDLTEQSGSGKSHLEKIRKQLEEK 1537
QY 1331 RIIVKEVELEEEVEKEVAKVGTENTLFQ-----RLAESFNELKODENFVBLFS- 1382
DB 1538 TELQSLAEAEASLEHEESKILRAQLEFNQIAELIERKLAEDMEDEQAKRNHLRVVDSL 1597
QY 1382 -----KEEME-FLDEIKMEASVEKLEFGKA 1405

```

```

DB 1598 QTSLEAETSRNEALVKKKMGDNLNEMETQLSHANRAAE 1639
RESULT 12
A:71928
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
C:Accession: A71928
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Voris, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557
A:Accession: A71928
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1819 <ARN>
A:Cross-references: GB:AE001481; GB:AE001439; NID:94155005; PID:94155007
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf13/14

Query Match 1.9%; Score 142.5; DB 2; Length 1819;
Best local similarity 19.0%; Pred. No. 1.1; 524; Indels 417; Gaps 68;
Matches 267; Conservative, 199; Mismatches

QY 178 LPEENVVOQFNECAVSRKKCVPRKSDVGEFPVPRNAVONFNKDFSGKWTYTSGLNPT 237
DB 244 LFSRSIFHYF-----VPLEDKSSRFS-KDRNLYVND----- 274
QY 238 FDAFDQOLHEFTHEMDKLVGNLTWRITLQDGEFTSRSAVOYTFVODPDLPALYNHNEFL 297
DB 274 -----EIQIRQ---EYNRLKERNKGNWIDKNLF-----FNDPN--RTLYN----- 312
QY 298 HYQDDWYLLSQIENKPPDIFYVYRGRNDAMPDGYGSVLYTSRPLPESILIPNLOKAAK 357
DB 312 -----YLNIAEIDK--NPLRAFYECISN---GGN--YEELKIKIN--KKLODOK 354
QY 358 SVGRDNENFITDNGSGPEPPIYERLKEATEBEGK--LLIKAVER-----EEVY 405
DB 355 KTLAYNDCT---KNAKTEEBRIKCIDLINDEMLKSLNQOYVALDCLKNAKTDEER 411
QY 406 EKEVEKVADEEM-TLFORLLEGFEKLOODEENFVRELSEKKEK-EIINLOMEATEVEKLE 463
DB 412 KECLKLINDPEIRKFPKLELOEYKDCIKNAKTEAKNECKLGSKEA--IERLK 469
QY 464 GRALPIKRLALAPHSNPLNHEHTIKYVGSKLPGRKRSWGWEDYFAGIVAVAKICSSR 523
DB 470 QQALDCLKNAKTDEERKECLKN--IPQDLQKELLADM-----SVKAYKDCVSR 515
QY 524 RIPRYFRK-----SPRICGLDSRGLOLFSHGKHLNSPAHSINQWPK-----GNSG 570
DB 516 ARNKEKQOBECKLITPAKKLLENQALDCLKNAKTDEERKECL-KNLPKLOSLDIKAKS 574
QY 571 CKFPKYDA-----LMWEKMGQFAKTAIVAIFILSVASKADAVDALKTCTC 616
DB 575 LKAYKDCASQATEAKEKECEKLLTPE-----AKLLEBAKSVKAYIDCVSQAT- EA 628
QY 617 LKRC-----RELAKCISNPACANVACLOTGNRPDETQO---IKGGLFE 662
DB 629 EKKECEKLLPEKAKKLEK---KSVRAYIDCVSKANNEERKECKLITPEAKKLE 684
QY 663 NSVADDEFNECAVSRKKCVPRKSDVGEFPVPRDPSVLQKFDMDQFGSKMPTITGLNPTDPA 722
DB 685 NQALD-----CLKNAKTDEERKECLKLP-----KDLOKKVLAESVRYIDC 727
QY 723 FDOQLHEFTHEMDKLVGNISWRIPTPDGFTSRSAVOKRVQDPKYPGILY----- 773
DB 728 VSRKKNNEARKECE-----KLTLP-----ARKLLEBAKSVKAYKDCVLRANE 772

```

QY 773 --NHNDNYLLQDDWYLLSSKVENSPEDYIFVYKGRNDAMDGYGGSVLYTRSAVLPESI 830
 Db 773 KEKOECEKLLTPPEARKLLEES--KRSVAYLDCVSRANKAEKECEKLLTPPEARKLLEEA 831
 QY 831 IPELQTAAGKVGDFNTFTKTJDNCTGPEPLVERLEKYEGEREPT-----IIEVEEPI 883
 Db 832 KESVAKYKOCVSARNE--KEKOECE--EKLLTPPEARKLLEESKSVAYLDCVSRANKA 887
 QY 884 E-EVEKVDKREYLLFSKLEFEKEL-----QDEENFLRELSE-----EM 924
 Db 888 ERKECEKLLTPPEA--KLLLEEKESVAKYKDCVSRANKKEKOECEKLLTPPEARKLLEN 944
 QY 925 DVLDDGLKMEYVE--KLFGKALP---IRKLMVATHCFTSPCHDRIRFPSSDDGIGRLGI 980
 Db 945 QALDDCKNAKTEAEKRCVYKDPKDLQKVA-----KESVAYLDCVSRANKA 994
 QY 981 TRKRINGTFLKLLPPIQASADLRTTGGSSRPLSAFRSGSKGIPDIVLPKSKNELKELT 1040
 Db 995 ERKECE-----KLLTP-----EARKLLEEKESVAKYKDCVSRANKA--KEKOECEKLL 1041
 QY 1041 APLLKLV-----GVLA-----CAFLVPSA-----DAWDLKTC-- 1071
 Db 1042 TPPEARKLLEGEVKKSVAYLDCVSRANKKEKOECEKLLTPPEARKLLENQALDCKNAKT 1101
 QY 1071 ACLLKGRLEL-----AKCIANPACANVACLOTNNRPDETCO-----IKGDLFENS- 1121
 Db 1102 EAEKRCVYKDPKDLQKVAKESVAYLDCVSRANKKEKOECEKLLTPPEARKLLEESK 1161
 QY 1121 -VVDEFECASVRKCKVPRKSDLGEPAPDPVLYVONFNISDNGKWYITSGINPTFPAF 1179
 Db 1162 KSVKAYLDC--VSRANKAEKKECEKLLTPPEARKLLE-----AKESVAY 1205
 QY 1180 -DQOLHEFTEGDKLVGNISWRLKTLDSGFTFRSAVQKVOQPNOPGVLYNNDNEYLYH 1238
 Db 1206 KDC-----VSRANKKEKOECEKLLTPPEARKLLE-----EKEKLLT 1226
 QY 1239 QDDWYLLSSKIENKPEDYIFVYKGRNDAMDGYGAVYTRSSVLPISLPE-----LEKA 1294
 Db 1227 PEARKLLEGEVKKSVAYLDCVSRANKA-----KEKOECEKLLTPPEARKLLEEK 1275
 QY 1295 AKSIGRPFSTFIRTDNCTGPEPLVERLEKYEGEREPT-----IIEVEEPI 1354
 Db 1276 RQQDKMIAKCKL--NADPDRBAIMKCLDGLSDEKLT--KYLQEAKEKAVLDCLTART 1331
 QY 1355 E-----MTLFOALAEF--NELKODEENFVRELSEKEL-----MFLDE----- 1391
 Db 1332 DEERKCONLYSDLIOEIIONKKAQNKN--QLSKTERLHQASECLDNDLDDPTDQAEIQ 1388
 QY 1391 -----IKMEASEVEKLF 1402
 Db 1389 CLEGLSDSERALLIGIKRQADDEVDRIT 1415
 RESULT 13
 C70319
 Nitrite reductase (NAD(P)H) large subunit - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence-revision 08-May-1998 #text-change 08-May-1998
 C:Accession: C70319
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V
 V.
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MIMD:9819666
 A:Accession: C70319
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1002 <AOF>
 A:Cross-references: GB:AE000678; NID:g2982921; PID:g2982926; GB:AE000657
 A:Experimental source: strain VF5
 C:Genetics:
 A:Gene: nirB

Query Match 1.9%; Score 141; DB 2; Length 1002;
 Best Local Similarity 20.1%; Pred. No. 0.59;
 Matches 192; Conservative 148; Mismatches 314; Indels 300; Gaps 54;
 QY 289 LYNDNEFLHYQDDWYLLSS-----QIENKPDYIFVYKGRNDAMDGYGGSVLYTR 340
 Db 90 LVTKDGEERY--DKAILATGSKPFIPPSIKGVKKGVFT--YRTADVEILDIYARSKR 146
 QY 341 SPTLPSTIIP--NLOKAKSVGRPFNNITTDNCGGPEPLVERLEKTAEP-----G 390
 Db 147 AVVIGGGLLG--EYTKALRDIGLEVLVHIIDLTL-----MEQDLKTADELKRLDEDNG 200
 QY 391 EKLIRKAEVEIEEV--EKEVEKVR-----DTEMTLPORLLEGEKELQDEENFRELSEK 444
 Db 201 IKVLK--KYTEEILGEKRAEGRPSDGBELADFIYATG--IRPNVEGVNSGLKV 254
 QY 445 EKEIL--NELOAEATEV-----EKLGRALPIRKLMAIAPSNFLANHEITKY 492
 Db 255 NKGIYVNDYLETSAVDIYAVGECIEHKGTYGLVAPIME--QVKCAHN--VVHNGEKY 310
 QY 493 VGS-----KLPGHKRPSWG--WEDYFGSIYAKICSSRRIRPRFRKS-----PRIC 537
 Db 311 TGLTYAMAKYAGVNLFSAGINEKDDGVAFIDNRSLS--YTKAVIRNKIYGTILY 367
 QY 538 GLDSRG-----LQFSGHKHNLN-----PAHSINQNPKNNSGCKFPKDYALMWEKMGOFAR 590
 Db 368 G-DVAGNNYLLDLIKSG--DISEERPFLIKHLPKGTGTVEELK----- 412
 QY 591 TAIYAIPLISAKADAVDALKTCTCLKECRLELACISNPACANVACLOTNNRPDE 650
 Db 412 -----NDIVCNNAVTKG--EIVCIKE-----GC-KTLEEIOER 443
 QY 651 TEOCIKCGDLFENSVDENFENCAVSRKCKVPRKSDVGFVPYDPSVLYOKFDMKDFSGKW 710
 Db 444 TKAISTGSCGLE--LVEEILKHVYKEK--PKR-----VNKLEV----- 478
 QY 711 FTRGLNPTFPAFDCQLHEFHT-----ENKLVGNLSW-----RIPT 747
 Db 478 -IKKELHPDLEFKRRLKLYSSEGELENIPEEDRV--RLKMYGIFYRKATGYFMVRLRV 535
 QY 748 PDGGE--FTRSAVQKVOQPKPIGILYNNDNEYLLQDDWYLLSSKVENSPEDYIFVYK 806
 Db 536 PNGRLSYEQAKVSHISEKFCGEVEITSROOL--QIRMI-----KLKDLPELT----- 583
 QY 807 RNDAMDGYGGSVLYT-----BSAVLPESIIPELOTAO-----KVG 842
 Db 583 --EALNRVGLSLQIGMDNVRNVTGDPITGLAEDSLITLRSQETINIFLGKKRYADLP 640
 QY 843 RDNFTIKTDNCTGPEPLVERL-----EKKVEGERETII-----KEVEIE 884
 Db 641 RKLNAV--LGSQTDICINALFNDVCFYLSKDGKLGFLNYLGKIGSGGPKKAIMD 695
 QY 885 EEEKVDKREY--TLFSKLFEGKELODEENFIRELSKEMVDLCKMEATEVEKLF 941
 Db 696 MEVEPYEVEVEYKALP--DIYSTFGRNRRNSKRNRYFLLOEW--GVERFEELERL 748
 QY 942 GRALPIRKLMA--ATHCFTSPCHDRIRFPSSDDGIGRLGITRKRINGTFLKILPPI----- 998
 Db 749 YKAIIP-----SKGIDLNVKNGRGGIILNR-----NGYAVCVVYPAKIK 789
 QY 998 -----OSADL--PTTGGSSRPLSAFRSGSKGIPDIVLPKSN-----ELKELT 1040
 Db 790 AKDFQIAELARKYGSRRL--LSYQNY-----IPNIPENLPLLGEELEFEKESYVS 842
 QY 1041 APLLKLV-----GVLAFLIYVSADAVDLKTCACLLKCCRLELAKCIANPACA 1091
 Db 843 SPTMHLIACAGSDTOSFVGIPIKKS--DAVRVAKYLSERKLIDIFYRMWSACA 894
 RESULT 14
 JC6009

surface-located membrane protein Imp3 precursor - Mycoplasma hominis (SGC3)
 C:Species: Mycoplasma hominis
 C:Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #ext_change 05-Dec-1998
 C:Accession: J06009
 R:Label: J. A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
 J. Bacteriol. 178, 2775-2784, 1996
 A:Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis Imp gene system
 A:Reference number: J06009; MID:96213016
 A:Accession: J06009
 A:Molecule type: DNA
 A:Residues: 111302 <LAD>
 A:Cross-references: EMBL:X95601; MID:g1197335; PID:e222344; PID:g1197336
 C:Gene: Imp3
 A:Gene: Imp3
 C:Genetic code: SGC3
 C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homology
 C:Keywords: duplication; membrane protein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>
 F:957-992/Domain: tetratricopeptide repeat homology <TT1>
 F:993-1026/Domain: tetratricopeptide repeat homology <TT2>
 F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>
 F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

Query Match 1.9%; Score 140.5; DB 2; Length 1302;
 Best local similarity 18.7%; Pred. No. 0.92;
 Matches 173; Conservative 123; Mismatches 312; Indels 315; Gaps 40;

145 KCLANPS-----CAAVACIQT-----CNRDETEECIK--CGDLFF--NSVDQENE 189
 445 KSVTSSNKSNDIIAANEALIALADANKAKDOVDANKSIRKQALNIDKANTLLPOLND 504
 190 CAVSRKKCPKSDVEEPVDRNAVONFNMKDFSGKYTTSGNPTFADACQDLHEH 249
 505 ---NDEIYKAKESLNAEITNAKAVNQDNMSMOSAK-----SSJDDKYTKIQNDLTFEN 557
 250 MENDKLVGNITRKITKLDGGFFTRSAVQTFVDPDLGALYHNDNEFLHYODWYILSSQ 309
 558 KKKDA-----KFKLEO-----TRKDIDNFTLMD-----VKNNPN-----YATLVKD 594
 310 IENKPPDYIFVYRGNAMDGGSVITRSPILPESTIPN-----LQ 353
 595 LNAKDKKSV-----TKSSNKSEIIAANDKQALDKAKYAKQID 636
 354 KAAKSVGRDFNNFTTDSNGEPPVLERLEKTAEGEKLIRKAEVIEEVEKREKVR 413
 637 EANKSIEQLSDSITNANO-----LNKL-----VDSDDIKAKTELISQELQASQELN 686
 414 -----DLEMTLFFORLLFGRLEQODEENFVRELSKEKETLNELOMEATEV 459
 687 LNNPISMSGAKESLDKAVETITKLETF--NKDKVAKFELEKTRKDI----- 733
 460 EKLFGRLPIRKLRMALPHSNFLNHETIKYVGSKLPGHKRFSWGEDYFGSIVAKI 519
 733 -----DEFINTKNTNPNY-----STLISEL 752
 520 GSSRIPIFRFKSPRICGGLDSGLQDFSHGHNLSPAHS--INQVPGNSGCKFPKQVA 578
 753 TSKR-----DSKNVTSNNSKSDIETANTTELKQALAKANDKQADNLA 796
 579 LAMWKGQFAKTALVALFIIISVASKADAYDALKTCTCLLEGRCLAKKCSNAPCAA-N 637
 797 KSTRQOLNNSISSA--NLLAKLIDKDNITQOAKF-----ELKEVOK--ANOAVASNN 846
 638 VACIOTCNRPDETCQI--KCGDLFENSVDVEFNCAVSRKKCVPRKSDVGF-----PV 691
 847 TASMSAKSSIDAKYTEITLKLETFNKKDKVAKFKLEQIRK-----DIDEIFINTKTN 899
 692 PDPSTLVOKFPMKDSGKWFITRGINPT-----FDAFDC----- 726
 900 PDYSTLISELTSKRD--KNSITNSNKSIDEIETANTTELKQALAKANDKQADNLAIRSTKE 958

QY 766 QLEHFEHEENKLVGNLSWRIRTPDGGFFTRSAVQKFOV-----D 764
 DB 959 QLNKSISANTLLAKLTDKDNTIOQ---AKTELEKEVQKANOAVASNTASMSAKSSLD 1015
 QY 765 PKYPI-----LYNHDN-----EYLLYQDDWYILSSKYE-----NSPDIYFV 802
 DB 1016 AKVETITKLETFNKKDKVDKRELEQGRKDIDEFINTKNTNPNYSTLISELTSKRD---- 1072
 QY 803 YKKGRNDAMDGGGSVLYTRSAVLPESII-----PELQTAQKVGDRNFTFKTNTG 856
 DB 1072 ---SKNSTNSNKSDEIETANTTELKQALAKANDKQADNLAIRSTKEQLNKSISANT-- 1127
 QY 857 PEPPLVERLEKKEVEGERTTIKEVEIEEVEKVR-----DKEVTLF 898
 DB 1127 -----LAKLIDKDNITQOAKTELEKEVQKANOAVASNTASMSAKSSLDFTVTEI 1178
 QY 899 SKLPEGFKEIQDDEENFLRELK 921
 DB 1179 TKKLETF--NKDKAEKFNELKK 1198

RESULT 15
 F64508
 hypochlorite protein MJ1672 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #ext_change 10-Oct-1997
 C:Accession: F64508
 R:Bolt, C.J.; White, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
 rson, J.D.; Sadow, P.W.; Hanne, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
 A:Reference number: A64300; MID:96337999
 A:Accession: F64508
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-800 <BU>
 A:Cross-references: GB:U67607; GB:L77117; MID:g1592244; PID:g1500575; TIGR:MJ1672
 A:Genetics:
 A:Map position: REV1656207-1653805

Query Match 1.9%; Score 140; DB 2; Length 800;
 Best local similarity 20.0%; Pred. No. 0.49;
 Matches 180; Conservative 96; Mismatches 228; Indels 394; Gaps 45;

QY 650 ETECQIKGDLFENSVDVEFNCAVSRKKCVPRKSDVGFPPDPSPV--LVQKF--DKMD 705
 DB 119 EYVC---IGELFENLYKNGFK---YSLKPLANSDAIFTDKPPPNENYKELPSKFEDEIKD 172
 QY 706 FSG-----KFTI-----TRGINTPFAFD-----COLHEFH 731
 DB 173 FKGDVFEELIYLMOKYTCIPSVTMMKAGSLKGLDPVSLFDSKTTCAIACCLXYMY 232
 QY 732 TEENKLVGNLSMRIRTPDGGFFTRSAVQKFOVDPYRGLYHNDHEV-----ILY----- 782
 DB 233 VKENK-----KKNKYKEYIDDKTLEKLFNNDGKMKELFSLIHGDS 275
 QY 782 ---QDDWYILSKYENSPEDIYIFVYKGRNDAMD-----GYGGS 817
 DB 276 GIDDFEFITTK-----YATKSLKGRSFYDLFTEYFAKICKELMLPTINILFYGG 328
 QY 818 VYTRSAVLPESIIPELQTAQKVGDRNFT-----IKDNTGCEPPPLVERLEKKEVEGE 873
 DB 329 HYILSYKVDENLKNKLEINEV--LFNMFRTKIYITIAEVGVTPNDPKFSKSDK--ESKE 385
 QY 874 RT--IIEVEEIEEVEKVRKDEVTLFKLEFG--FKELQDDEEN----- 915
 DB 386 KTWGFAKKWKEVS---EKVTEKKLRFFYKLEGLFEPNRSSENNKVCICRNEFDKNEGY 442
 QY 915 FLRE--LSKE-----MDVLDGKKMEATEVEKLFGRALPIRKLRMAVATHTCTSP 961


```

Db 443 AIRENESSEKICDYCASVALTDILKNFQMEKT---IKFNKAYPI----- 486
QY 962 CHDRIREFSSDDGIGRIGITRRKIRINGIFLLKILPPIQADIRITGGRSSRPLSAFRSGFS 1021
Db 486 -----IHUTKNKDN-----LSLQREFFS 503
QY 1022 KGIPIVLPSPKNEKELTALPLKLVGLACAFIIVPSADAVDALCTCACILKGCRIEL 1081
Db 504 -----FLTV---KALELESKFRVLSD----- 523
QY 1082 AKCIANPACANVACLQTCNNRPDETEQIKCGDLFENSVDDEFNECAVSRRKCVPRKSD 1141
Db 523 -----ENYFLKEKKLPDSCGELIIPYKIM 546
QY 1142 LGEFPAPDPSPVLVON---FNISDENG---KWIYITSG-----LNPTDPAEDCQLHEFHTE 1189
Db 547 AIAFP-----IENETEKRIIDFDGLAEKAFERTGTGRKIGILKMDVD---NLGEIIFTT 596
QY 1190 GDNKILVGN---ISWRIKILDSG---FTTSAVQKFVQDPNPGVLYNHNDNEXLHYQDDWY 1243
Db 597 G---LGNDATIS--RMSLSSMLTLEFTGYI-----PHLIKNEEFE----- 633
QY 1244 IISKIEKKPEDYIFVYRGRND---AMDYGGAVVYTRSSVLPNSIIPLEKRAKSI 1298
Db 633 -VNGK-KYKFKONIYLYVAGGDTLIYGAMDA-----VWELAKRI 670
QY 1299 GHDESTFIRTDNTC-GPEPAL-----VERIEKVEEGEIRIIVKVEEIE 1341
Db 671 RGDFFKFFV---CYNPYITLSAGIVFVNPKEEFKRAVMAEELENGKNYIIEDEDETE 725
QY 1342 EYVEKEVEKVGRTM-----TLQRLAEGFNEFKODEENFVRELSKEEME 1386
Db 726 KKVDNALIVNCPNMNMLEVYNEYCWTKLKSYLEGINKEMVLESLVKKFNEDDLE 783

```

Search completed: October 14, 1999, 03:19:31
 Job time: 2217 sec

•
•
•
•

OM protein - protein search, using sw model

Title:	US-09-075-375-4
Perfect score:	7495
Sequence:	1 MASLSHTVFLCKEEALNLYA.....MEASEYEKLFGRALPIRRVR 1412

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	158.5	2.1	1957	1	YD86_SCHPO	Q10411 schizosacch
2	155.5	2.1	2868	1	RBPI_PLAVB	Q00798 plasmodium
3	155	2.0	1	TEGU_HSUSA	001056 hepervirus	
4	148.5	2.0	1935	1	MYSP_PIG	P79293 sus scrofa
5	146	1.9	1935	1	MYSM_HUMAN	012883 homo sapien
6	143	1.9	2748	1	NUM1_YEAST	Q00402 saccharomyc
7	142.5	1.9	1935	1	MYSB_RAT	002564 rattus norv
8	137.5	1.8	1875	1	MLPI_YEAST	002455 saccharomyc
9	134.5	1.8	3660	1	MDP_CHICK	PI1533 gallus gall
10	134.5	1.8	1937	1	MYSP_HUMAN	PI1533 homo sapien
11	134	1.8	1630	1	MSP1_PLAFK	P04932 plasmodium
12	134	1.8	1639	1	MSP1_PLAFW	P04933 plasmodium
13	133.5	1.8	1934	1	MYSP_MESAU	PI3540 mesocricetu
14	132	1.8	809	1	LEF_BACAN	PI1597 bacillus an
15	130	1.7	1679	1	Y109_YEAST	P40457 saccharomyc
16	129.5	1.7	1805	1	HMW2_MGCE	P47460 mycoplasma
17	129	1.7	1822	1	YM68_CAREL	P34529 caenorhabdi
18	127.5	1.7	1790	1	USO1_YEAST	P25386 saccharomyc
19	126.5	1.7	4568	1	DYH6_CHIRE	Q39365 chlamydomon
20	126	1.7	993	1	NISB_LACILA	P20103 lactococcu
21	126	1.7	1744	1	TANA_XENIA	001550 xenopus lae
22	124	1.7	2284	1	POLI_GFLV	P29149 grapevine f
23	123	1.6	3678	1	DMD_MOUSE	PI1531 mus musculu
24	123.5	1.6	1341	1	RPAL_MERJA	058357 methanococc
25	121.5	1.6	1039	1	STL_MERJA	058357 methanococc
26	121.5	1.6	1545	1	TMH2_YEAST	Q03640 saccharomyc
27	121	1.6	2025	1	TTCC_HUMAN	P53804 homo sapien
28	120.5	1.6	2418	1	SPCA_HUMAN	P02549 homo sapien
29	120	1.6	610	1	SCGA_HUMAN	014967 homo sapien
30	120	1.6	1324	1	CUT3_SCHPO	P41003 schizosacch
31	120	1.6	4725	1	DYHC_DICDI	P43026 dictyostellei
32	119	1.6	881	1	ARBP_YEAST	012386 saccharomyc
33	119	1.6	2411	1	MYSA_DROME	P05661 drosophila
34	119	1.6	709	1	TOP1_MYGE	P47368 mycoplasma
35	119	1.6	1102	1	YM65_YEAST	Q03660 saccharomyc
36	118	1.6	2663	1	CENE_HUMAN	Q02224 homo sapien
37	117.5	1.6	857	1	PIF1_YEAST	P07271 saccharomyc
38	117.5	1.6	1230	1	UGS4_SOLU	P43846 solanum tubu
39	117	1.6	2095	1	RRPL_TOSV	P37800 toscana vi
40	117	1.6	2363	1	SPCO_MOUSE	062261 mus musculu
41	116.5	1.6	1940	1	MYSE_RAT	PI2847 rattus norv
42	116.5	1.6	1208	1	YDH6_SCHPO	092351 schizosacch
43	116.5	1.6	823	1	YHX2_YEAST	P38683 saccharomyc

[illegible]

```

Db 641 QOLERANPEKESTLSDENNDLRKLKLEESNKLKQEDVDSLEKNOTL----- 696
QY 496 KLPGRHREWSMEDYRGSIVAKICSSRRIPRYFRKSPRICGGLDSRLQLEF---SHGKH 552
Db 696 -----ED-----LRKSEALRFSK-----LEAKNLREVINDLKGKH 726
QY 553 NLSPA-----HSI-----NONVPKNGSGCKPKPDVALMWKMGQFATAIATVILLSVA 602
Db 727 ELLERQNRDLHSLSPAKNTNMLISSELTSSSEDV-----KRLANWETTLQD 774
QY 603 SKA-----DAVALKTCTCLKECRLELAKCISNPACANVACLOTNNRPDETECOIK 656
Db 775 SKAMROSEFTSLVNSYSQISLNYELHEDHVN-----MQSQNNLLSESESLK 821
QY 657 ---CGDLFEN---SVDEF-----NECAVSRKCCVPRK-----SDVGEFVDPDP 694
Db 822 TPCENITQONMFLIDNVOKIMKHVNOESKVSSEKVSGLSLDKLNRSSLVNAISDND 881
QY 695 SVLVQKFDMA-KDPSGKWFITRGLNPTPDAPDQLEHFEHTEE-----NKLGNLSMRIR 746
Db 882 QLTQLELSTKRYDLEQESAGLNGSLKSLERKQLHTENDELHRLDKLTGKLK---- 938
QY 747 TPDGFFTRSAVQKRVQPKRYPGILYNDNEYLLYQDDWYLLSKVNSPEDYLFVYKKG 806
Db 938 -----IESSKSSDL-----GKKLIARQEB---ISNLKEENMSQQAITSVKS 976
QY 807 RNDAMDGYGSVLYTRSAVLPESITPELOTAAOKYGRDPENFIFTDNCGEPEPLVERLE 866
Db 977 KLDE-----TLSSKSKL---EADLEHLKKNVSEVEVRNALLSN-----ERLM 1017
QY 867 KAVF-EGERTIKKEVEELIEEVEKVRKQEVTLFSKL-----PDGF-----KELORD 911
Db 1018 DDLKNGE-----NIASQTEIEKKRAENDLOSRLSVSSSEYENLLISQTKSLSE-D 1071
QY 912 EENFRELSEKEMDYLQDKMEATEVEKL---FGR-----ALPIRLKMAVATHC 957
Db 1072 KTNOLKYTEKVNQKLLDKDORVLEELISKYGLKEENAOIKDELIALRKK----- 1125
QY 958 FTSPCHDRIRFESSD-----DGIGRL-----GITRRKINGTFLIKILPIOS--AD 1001
Db 1125 -SKQHDICANFVDDDLKESDALQOLETNEKVELVLSLEQSSNNNEALVERSDLANLSD 1183
QY 1002 LRTTGGRSRRLSAFRSGFSKGTIDYVLPBSKNELUJAPLLIKLVGLACAFILVPSA 1061
Db 1184 MKKSLSDSDNIVSVIRSD-----LVRVN 1206
QY 1062 DAVALKTACLLKGCRLELAKCIANPACANVACLOTNNRPDETECOIKCGDLFENSV 1121
Db 1207 DELDTLK-----KDKDSLSTOYSEVQDRODLDLSLKGK----- 1241
QY 1122 VDEFNECAVS-RKRCVPRKSDLGEPAPDPSVLVN--FENISDFNGKWYIT-SGLNPTFD 1177
Db 1241 EESFNKAYVASRELCTKSEIDV-----FVSEIIDDNFVFNAGNSSELSRLITVLSLEYVD 1295
QY 1178 AFDQLEHFEHTEGNGKLVGNISWIKITLDSGFFTRSAVQCFVDDPNQGVLYVHNDDEYLH 1237
Db 1295 AFN-QVAFKKMELDN-----RLTTTDAEF-----TRVVDL-----EKLOH 1330
QY 1238 YODDWYLLSKIEKRPEDYLFVYVGRNDAMDGYGAVVYTRSSVLPNSTIPELEKAAS 1297
Db 1331 EHDOWL-----QRG-----DLERAKLD 1348
QY 1298 IGRDFEFTITDNTGEPALVERIEKIVEEGRIIVKEVEELIEEVE----- 1346
Db 1349 SEKVF-----LRKEAMTENIH-SLEEGKEETKKETJAEILSSRLDNOQLATNKLNG 1398
QY 1346 -----KEV-----EKVGRTEMLTFQRIAGFNLKODENFVRELKSKDEML--DEIKMA 1395
Db 1399 LDHNGQIRLKEVDYKESLITLESJUSNOROKESLTD--AKNDELHMDDTSRKNS 1456
QY 1396 SEVEKL 1401
Db 1457 SLMEKI 1462

```

```

RESULT 2
RBP1_PLAVB
ID RBP1_PLAVB STANDARD: PRT: 2869 AA.
AC 000798;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1.
OS PLASMODIUM VIVAX (STRAIN BELEM).
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN (1)
RN SEQUENCE FROM N.A.
RX MEDLINE; 9231538.
RA GALINSKI M.R., MEDINA C.C., INGRAVALLO P., BARNWELL J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL CELL 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M88097; G160626; -
DR HSP; P36956; IAN9.
KW MALARIA; RECEPTOR; SIGNAL; TRANSMEMBRANE.
FT SIGNAL 1 17
FT CHAIN 18 2869
FT DOMAIN 18 2807
FT TRANSMEM 2808 2826
FT DOMAIN 2827 2869
FT SITE 1030 1032
FT SITE 2599 2601
FT SITE CELL ATTACHMENT SITE (POTENTIAL).
FT SITE CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; 2BD43E2C CRC32;

Query Match 2.1%; Score 155.5; DB 1; Length 2869;
Best Local Similarity 17.4%; Pred. No. 0.59;
Matches 283; Conservative 242; Mismatches 573; Indels 529; Gaps 76;

QY 57 TSKYSSSDSHCKDKQISDTSEET-----QRFDLKRGMTLILKQWRQFTQLA 110
Db 1150 TSNKEIILSKYKEVEDKUNLVEQNEDEYRKVKNPENEOLEAIRISMSKLE----- 1201
QY 111 IVLVCTFYVPRVDAVDALKTACCLKECRLELAKCIANPACANVACLOTNNRPDETE 170
Db 1201 -----VINKHVEEMQLESTANTLK-----SNAKKENENDLELN---KTK 1239
QY 171 COIKCGDLFE--NSVDOFNECAVSRKRCVPRKSDVGEFVDPDRNA---VVQNFNM-KDF 224
Db 1240 GQMR--DIYEKILKRIAELEKGTIVNELKANDNEKANKVE-PEPERNIIGHLEIRIYEKK 1296
QY 225 SGKWTYTSGLNPTPDAPDQLEHFE--HMENDKLVGNLTWIKITLDSGFFTRSAVQCFVDD 282
Db 1297 AGK--VDEKMSLTKTIEKQLQETSDDSONELVYTSITKLELNAKG----- 1341
QY 283 PDLGALYNHDEFLAYQDDWYLLSQIENKPPDYLFVYVGRN-DAMDGYGSVLYTRSS 341
Db 1341 -----YED-----VIRKNEDSITOLREKAKASLETLD-----EM 1368
QY 342 PTLPESTIPNLQRAK---SVGRDFNN-----FITDINS----- 373
Db 1369 KRLVQVQNMNLQSAIQNAGISRELNLKGVIELLISTNYSILEYVKKNSSEVFSQL 1428

```

```

QY 373 CGPEPPLVERLEKTA---EGEKLLIKEAV-----EIEEVEK-----EVEKVRDT 415
D 1429 ANGETTKAEKEKNASARLAIAEKL---KEQIVKDLDYSDIDKVKKLEGIKREILMKKS 1486
QY 416 EMTLFORELEGGKQOQOENFVRELSEKEEILN-----ELOME-----456
D 1487 ALTFEEF---SEKFKQKSSHMENAKEGKKIYILKNNNGDGKANTDQOMEVENVYSKAE 1545
QY 456 -----ATEVEKLEGRALPIRLKLMALAPHSNFLANHETIKY 492
D 1546 HAFHTEVAQVKTAFCEISIVAVYTKMDNLFENSL-MKEVYVCKEK-----KNDKAEKY- 1599
QY 493 VGSKLPGHKRSWGWEDFGSIVAKICSSKRIRPRYKRSRICCGDSGLQFSGKH 552
D 1599 -SAKL-----KPYDGR-KARVSENERKISELKEKAV---BKSSQLNDVSTK 1643
QY 553 NLSPAHSINONVPGKSGCKFPKVALMWKMGQFAKTAIVAFILSVASKADAVALK 612
D 1644 SLIQDNCROOLDSTVLSNIGRKONALOYFDS---ADSKSVSLPISELGAEKSLDKVK 1699
QY 613 TCTCLLECRLELARCISNPACANV---ACLOTCNNRPDETECO-ICGDLFENSVDDE 669
D 1700 AA-----KESEKNELETVONEMSRINVEEGSLTDIDKTTIENDLLMKKQYEGGLQKI 1755
QY 670 NECAVSRKKCVPR-KSDVODFVPPSVLVQKFDKDFSGKWFITRGLNPTFEDAFDQOLH 728
D 1756 KENAKRKRNSFELVGESEINALLDPSTSLFI-KLKLEVD-----MTGDLKMYGVKMN 1806
QY 729 EFHTEENKLVGNLSWRI-----RTPDGGFTFSAVOK 760
D 1807 EIHGETKSYNLIENHLSNATYVTFEKAQSLRELAKEEHLLRRREEMAFILNDIKK 1866
QY 761 FVO-----DPKYPGILYNHNDNEYLYVODDWYLS-----790
D 1867 VESLKLKEMMKKVAEYEGMKRDTSVQSLVODMKTTIYDELKTLNDISECSSVLNNVVS 1926
QY 790 --SKVENSPE-DYIFYYGGRNDAMDGGVGLYTRSAVLPESSILPELOTAQVGRDN 846
D 1927 IYKAVESKHADT-----RDANSMTESVTLNATYLSDEAKISS-----GMEFN 1971
QY 847 TFIKTD-----NTCGPEPPLVERLE-----KKEVEGER-----TIKEV 880
D 1972 AEKMSFKTDLELIFSVSINSNELLKITEQNSNDVIOKERSSEQLAKADIDYAVIKLK 2031
QY 881 EETEEVEKVRDEKVTLEKLEFGEKELORDE-----ENFIRELSKEDMDVLDLKRHEA 934
D 2032 NEFNEKLEAKKNEEYVASEKVRALKRLSQVEGIRCHENFRLDNT-----2081
QY 935 TEVEKLEGRALPIRLKMAVAHCTSPCHDRIRFESSDDGIGRLGITRKRINGFLIKIL 994
D 2081 -ELENL-----KKWYTI-----RDKKSREGLODM---EKENM-TYSNFI- 2118
QY 995 PIQSDALRTTGGRSSRPSAFRSGSKGIFDIPLPSKNE---LKEITAPLLIKLVGV 1050
D 2118 --TQLEGIIVASAGESKE-----DIEKLEHSEMEMRINISEKISTIDSK--- 2158
QY 1051 LACAFILVPSADAVALKTCACILKCRIFELAKCIANPACANVACLOTCNNRPDETECO 1110
D 2158 -----VIENNSTIDELYK---GKNQCAHWSILSYTANMKTSSKLMINKEKNT----- 2207
QY 1111 IKGDLFE---NSVDFENECVAVSRKKCVPRKSDLG-EFAPDPVSVLVQNFNSDENGWY 1167
D 2207 -KCYDVIKDNSSSTIDGVEL-----KGFYSGKLFFSSASELVQADIVSNPFAKH 2256
QY 1168 ITSGINPTFADQOLHEHFTBEDNKLKVNISWRITLDSGFTTSAAVQKFOVDNPQGV 1227
D 2257 EKESLNAIRD-IKKELYLFHONSIDISIV-----EGGYONML-----A 2292
QY 1228 LYNHNDNYLHODWYLLSKIEKKPEDY-----IFYYRGGRDA-----1268
D 2293 LYDKLNEKREMDLYLNISETKLQOMEHSTDVFKPIELHKGMETNNKSLLEKELKL 2352

```

```

QY 1268 -----WDGYGAVYTRSSV--LPPN--SIPELEKAASIGRD-----1302
D 2353 SVNDHMSMEAFKTKNKLKTPESVQNNINNVISIAEAKTLEIDROVDGNYQIVEEHK 2412
QY 1302 --FSTFIRTDNTGCGEPALVERIEKVEEGE---RIIYEEVEIEEVEKEVEK--GR 1353
D 2413 KQFSLIDRTN-----LMDIDIEFKKENNNYMLMEVNFETHRVNDYIEKTKNLVQAK 2466
QY 1354 TE-----MILFOR---LAEGFNLKODENFVRELSEKEM-----EFLDE 1390
D 2467 TEYQOLEENKQNDMLQNLFLKVSITIEFENVKRRKESLNDLYEQERLLKIGEHDE 2526
QY 1391 IKMEASE 1397
D 2527 IKRANTE 2533

RESULT 3
TEGU_HSVSA
ID TEGU_HSVSA STANDARD; PRT: 2469 AA.
AC 001056;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE PROBABLE LARGE TEGUMENT PROTEIN.
GN 64 OR BERE2.
OS HERPESVIRUS SAIMIRI (STRAIN 11).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 9233688.
RA ALBRECHT J.-C., NICHOLAS J., BILLER D., CAMERON K.R., BIESINGER B.,
RA NEWMAN C., WITTMANN S., CRAXTON M.A., COLEMAN H., FLECKENSTEIN B.,
RA HONESS R.W.;
RT "Primary structure of the herpesvirus saimiri genome."
RL J. VIROL. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92230228.
RA NICHOLAS J., CAMERON K.R., COLEMAN H., NEWMAN C., HONESS R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT organization between HVS and Epstein-Barr virus."
RL VIROLOGY 188:296-310(1992).
CC - FUNCTION: TEGUMENT PROTEIN.
CC - SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X64346; G60385;
CC DR EMBL: M86409; G330993;
CC DR PIR: H35812; H35812.
CC SEQUENCE 2469 AA: 280165 MW: 1872559 CRG32;

Query Match 2.1%; Score 155; DB 1; Length 2469;
Best Local Similarity 17.3%; Pred. No. 0.51;
Matches 230; Conservative 208; Mismatches 497; Indels 396; Gaps 58;
QY 271 FTSSAVQTVQDPDLPGALYNHNDNEFLHYODDWYLLSQIENKRPDYIFYYR-----324
D 178 FSTSNVDAIEYLSPPNVOYT--GSFLYFVPRKEYIGHSH-----YIMNYRVINTEKL 228
QY 324 -GRNDAMDVGGSVYI-----TRSTPELSIIPNLQKAKSVGHDNNFI 367

```

Db 229 HGNIDLTSGEGLIIEISPPNTPKPTSTOKPPTPTPKPAKPKTPRKKTREKSTI 288
 QY 368 TTNSGCPPEPLVERLEKTAEBGEKLLIKAV-----EIEEVEKEKVDLT--EMTL 419
 Db 289 PYDKS--KKRP--KIPKTSKSKSVLTKRDTALTPQKRIIEEHLRELLPPIREYEDMTL 343
 QY 420 FOR-----LIEGKRELQ-----QDEENPRELSKEKELINELOMEATEVEKL 462
 Db 344 FNNPVERTPGIDSLSGINSTTKREDDEDONVTSKKEDEDMWD----- 393
 QY 463 FGRALPIRLMALAPHSNFLANHETIKYVGSKLPGHRRFSWGEDYFGSIVA----- 518
 Db 393 ---IPPEVLDTETHS---DOETI-YMIGDE--NIMDMYSDDDDIDTIDISFIOLD 441
 QY 518 -KICSSRRIPRYFRKSPRICCGDLSGLOLFSHGK-----HNLSPASINONVPG 567
 Db 442 NLITSLDNIPKN-NTPPRI--IDKTSNPIKEGKALHSIDLKLNIVLEHGLITSSIS 497
 QY 568 NSGCKEPKDVALLMWEKMGOFKTAIVAFIISVASKADAVALKTCCLKECELELAK 627
 Db 498 ISKCSLQFVL---WGE-----KLSIPTR-DKLTITTELITELIEMIALTK 542
 QY 628 CJSNPACANAVACLOTNNRPDETCQIKGDLFENSVADEFNECAVSRKCVPRKSDVG 687
 Db 543 -LTN-----DTERNNVITKLKCMLEK-----SESVD 569
 QY 688 DPP-----VPDPSVLQKFDN-----KDFS-----GKNWITGLNPTF 720
 Db 570 SYKHLISALLNNIILKIQIIDTEIELKISTVFTSELGKDFSVCTKKESETIIMAKNLK 629
 QY 721 DAFDCQLEHFEHTENKLVGNISWRIRTPDGGFFTRSAVQK-----VODPKYPG 769
 Db 630 EKISTRKOELEENYFQSVL-----IAMETQPIPLPTRIEIOPSKKAQ 675
 QY 770 ILYNNDNYLYQDDWYLLSKVEVSPDYIFYVYKGN-----AMDGGSVLYTRSAVL 826
 Db 676 QL--HEKSLVEOK--LITIDANNVLDLHKMOKDITDISPPD-----PTVLKN 722
 QY 827 PESIPELOIAOKVGRD--FNFTIKTDNCGPEPLVERLEKKEVEGERETI--IKEVEE 882
 Db 723 IOSTLOLQVCTVDLNDIKKFIISNVQOLSTIGWEVALHSISQNNFRKADVLYLKLDD 782
 QY 883 IEEVEKVRDEEVLTFSLFGFGEKELORDEENFRELISKEEMVDLGLKMEATEVEKLF 942
 Db 783 IKKELOOVYTKO--KNEETLSKILADVOTLLENAKOSDLSIPILQHYITKAGTLVG 837
 QY 943 RALPIRLMAVAHCFETSPCHDRIRFESSDDGIGRLGI--KRRINGFTLKIPIPIQ 998
 Db 838 E-----RENOKFESLKYTVOKLSTSEELKTLIDSTLENVQLOIQ 878
 QY 999 S-ADIRITGSSRPLSAFRSGFSKG--IFDIYLPFSKNELEKELTAPLL----- 1046
 Db 879 EISDLOSNOYIHOSETIKQAFEPKSNIIINNIIQLOKQYTTVTPMLIAVKRELSEA 938
 QY 1046 -----KLVGLACAFLLVPSADAVDAKTCACILKCRITLACIANPACANVAC 1096
 Db 939 KFRRESNTICEITITLVSIGLSKSTVEAKDALKSIDLKEKL----- 984
 QY 1097 LQTCNNRPDETE-----COIKCGDLFENSVADEFNECAVSRKCVPRKSDGEEPPAP 1149
 Db 984 --TAVDRPLKRELTVNIRLQKLTLEQGEFDN-----KMEVDSF--VPI 1027
 QY 1150 PSVLVQNF--NISDFNKKWITISGLNPTFAFDCQLHEFTEGDN-KLVGNISWRITKID 1206
 Db 1028 PSRDKVKTFIQVAPSMAKQYAKKALKDOIQAMEIDVDPESVIEDNTKANQKAMQ----- 1083
 QY 1207 SGEFTRSAVQKFPVDPNQGVLYNHNDNEYLYQDDWYLLSKLENKPEYIF-----VY 1261
 Db 1083 -----KISAFQDLNFSILI-----PDDWLST-AKEYTRKSLTFLVIGPITLL 1124
 QY 1262 RGRDAMDGQGGAVYTRSSVLPNSIT-----PELEKRAK 1296
 Db 1125 KVEEVLVSYNLKEAKLKLSPNGVFTPPKFDWIIHYESNVNFHLKTIINLPKSVTAH 1184

QY 1297 SIGRDFSTFIRIONT--CGPEPALVERLE-----KTVF-----EGERIVKEV 1337
 Db 1185 NIGHELSSLQALNSKTLPPAVVGTISIEOHANAFSCMEFTLEATMHDVDRTRIDEYI 1244
 QY 1338 EEIEEVEYK-----EVEKYGR-----TEMFLFORLAGFENELKODEENFVRELISKEEMEFL 1388
 Db 1245 EDLRNDTKRHIVAPQIOSPPRFLSPEDIOEINSLPRLFDSILNENSRLLASQKNFOWL 1304
 QY 1389 DEIKMEASEVE 1399
 Db 1305 -ENTVKAHEIQ 1314

RESULT 4
 MYSB_PIG STANDARD: PRT: 1935 AA.
 AC P79293;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
 GN MYH7.
 OS SUS SCROFA (PIG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-DOMESTICA;
 RA KO Y.L.;
 RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC - FUNCTION: MUSCLE CONTRACTION.
 CC - SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC - SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC - DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC - PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC - EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMN)
 CC AND 1 HEAVY MEROMYOSIN (HMN). IT CAN LATER BE SPLIT FURTHER INTO
 CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
 CC - THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE MYOSIN, WHILE THE
 CC BETA ISOFORM IS A 'SLOW' ATPASE.
 CC - SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: U75316; GI698895;
 DR PFAM: PF00063; myosin_head; 1.
 DR PFAM: PF00612; IQ; 1.
 KM MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;
 KM ATP-BINDING; METHYLATION; ALKYLATION; MULTIGENE FAMILY.
 FT DOMAIN 1 838 GLOBULAR HEAD (S1).
 FT DOMAIN 839 1935 ROD-LIKE TAIL (S2 AND LMN DOMAINS).
 FT DOMAIN 780 802
 FT DOMAIN 839 1935
 FT NP_BIND 178 185
 FT DOMAIN 655 677 COILED COIL (POTENTIAL).
 FT DOMAIN 757 771 ATP (POTENTIAL).
 FT MOD_RES 129 129 ACTIN-BINDING.
 FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 695 695 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 705 705 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1935 AA: 223109 MW: F072F29E CRC32:

Query Match 2.08; Score 148.5; DB 1; Length 1935;
 Best Local Similarity 16.78; Pred. No. 0.87;
 Matches 248; Conservative 220; Mismatches 555; Indels 459; Gaps 57;

QY 166 PDETCQKCGDLPENSVVDQFNCAVSRKCVPRKSDVCEFPVDPDN-----AVY 216
 DB 375 PDGTE-----EADSAVLMGLNSADLLGLCHPRKVGKNGNYTGGQVNOCVATGALTA 428
 QY 217 QNFNKKDFSGKWTGSGINPT-----PDAPDCQJHE--FHMENDKL 255
 DB 429 KAVYEKMN--WMTYR--INTTLETKQPROPIGVLDIAGFEIPEFNGFEDCINFTNEKL 485
 QY 256 -----VGNLTWRKTLIDGCFTRSAVQTF-----V 280
 DB 486 QQFNHMFVLEOEYKKEGIEWEP--IDFGMDLQACIDLLIEKPMGIMSLIEECMFKA 543
 QY 281 QDPDLPGALY--NHONEFLHYODDWYILSSQJENKPD--DYIFVYRGRND-----AMDYGG 334
 DB 544 TMTFRKAKLYDNHLCKSNFQK-----PRNTKGRPEAHFALIHAGIVDVIIGM----- 594
 QY 335 SVIYTRSPITPESIIPINLQKAA--KSVGRDFNNFIITDNSCGPEPPLVERLEKTAEESEKL 393
 DB 594 --LQKKKDPINTVYDLYKSSKLKLSNLFANY-----AGADTP--VEKGGKAKKGSSE 644
 QY 394 LKAEVLEIEVEVEKVEKVRDTMTLFFORLLEGFELOODEENVRLEKKEKE----- 448
 DB 645 QTVSALH--RENINKLMTNLRST-----HPHFVRCIIPNETKSPGVID 685
 QY 448 ---LNLQELMEAT--EVEKELFGRALP-----IRKLMLALAHSPNFANHEITIKY 492
 DB 686 NPLVMQLKCNGLVBEIRICRKGFPNRIIYGFROGRYILNPAALPEQOIFDSKGAEXL 745
 QY 493 VGSKLPGHRRFSGWEDYGSIVAKICSSRRIPRYEPKSPRICGGLDSRGLQJFSHGK 552
 DB 746 IGLSLDHNQYKFGHTKVPFKAGLLGLLEMRDEKRLIRITRI--QAOSRVLSSMEFK 803
 QY 553 NLSPAHS---INQNVKNGSGCKFRPDVALWMEKGGQAKTAIYATILVASKADAVD 609
 DB 804 LTRRDSLLIOWNIR-----AFMSYKNN--PMKLYFKIKPLSAETEKEMA 850
 QY 610 ALKTCYLLKEC-----RLELAKCISNPACANVALCOTNNRPDETCQIKCGDLE 662
 DB 851 TMEEEGRLEKALEKSEARKELEKMSVSLQEKNDLQVQAEODNLSDEERDQJLIK 910
 QY 663 NSV-----VDENECVSRKK-----CVRKSDVDGFPVPPDSVLVQKFD 702
 DB 911 NKIQLEAKVEMTERLEDEENNAELTAKRNYEDECSELKRDIDLETLAKYERK-- 969
 QY 703 MKDFSGKWTTRGINTPTFAFDQQLHEPTEENKLVGNSWRIRTPDGGFTFRSAVQKRV 762
 DB 969 -----HATEK--VKNLTEKAGLD-----ELIATKL 993
 QY 763 QDPKPIGLYHNHDEYLYODDWYILSKVNSPEDYIFVYKGRNDADYGGSVLYTR 822
 DB 994 KKKALQEAHQALDLOAEDEKVTNLTAKAVKLEQHV-----DLAGSLDEOK 1042
 QY 823 SA-----VLPESITPELOQAQVVG-----RDF-----NTFKTNTGCP 857
 DB 1043 KVRMDLERARKLEGDKLQESIMDENKQDLERLKKKDELNALNARIEQALQS 1102
 QY 858 E-----PPLVERLEKKEVEGERT-----IIEVELEIEVEK----- 890
 DB 1103 QLOKKLKLQARIELEEL--EAERITARAVEKLRSLSELEIEISERLEAGGATSVQI 1161
 QY 890 -----VROKEVTLFSKLEFGKKEIQROBEENTLRRL 919
 DB 1162 EMNKKREAFQMRRLDEEATLQHEATPAALRKKHADSVALEGQIDNQKQKLEKK 1221
 QY 920 SKEMDVLDGKMEATEVEKLFGRALPIRKLMVATHCFSPCHDRIFRFFSSDDGIGRIG 979
 DB 1222 SEFKLELD---VTSNNEQITRAKANLEKM-----CRT--LEDQMEHRS-----KAF 1264

QY 980 ITRKRINGTFLKILTPISQADLRITGSSRPL---SAFRSGFSKGFIDVPLPSKNEL 1036
 DB 1265 ETQSVNDL-----TSQAKIQTENGELSRLDEKELIISQILRG-----KLVTQQL 1312
 QY 1037 KELTAPLLKILGVLAQFLIVPSADAVDAKTCACLLKGRCTELAKCIANPACANNAAC 1096
 DB 1313 EDLRQLEEEYKAKNALALAHQASARHAAADLRQNEYETEETAKELQRYLSK--ANSEYAO 1370
 QY 1097 LQTCNNRPDETECQIKGDLFENSVDDEFNECAVSRKCVPRKSDGFEPPADPVLVQN 1156
 DB 1371 WRT-----KRTDAIQRTELEPAKKLAQRLODAE----- 1403
 QY 1157 FNISDFNGKWTITGSLNPTDAFDQQLHEPTEENKLVGNSWRIRITLIDSGFTFRSAVO 1216
 DB 1403 -----AVANVAKCSLEKTKHRLQNEJEDLMDV-----VERSNNA 1438
 QY 1217 KVVQDPNPGVLYVNHDEYLYHODDWYILSSKLEKNPEDYIFVYRGRNDADYGGANV 1276
 DB 1439 AALDKKO---RNEDKILAEWKQKYEESQSELESSQK-----ARSL 1477
 QY 1277 YTRSSVLENSIIPLE-----KAASIGRDFSFIRDTGCPEDALVERIEKVEEGE 1330
 DB 1478 STELFKLNAHLSLEHLETSKREKKNLOEISDLTQOLGSSGKTHLEKVRKQLEAK 1537
 QY 1331 RIIVAEVEIEFEVEKEVEKVRTEMTLFO-----RLAEGFNEKODEENFVRLS-- 1382
 DB 1538 LELQALLESFASLSHEBQKILRAQLEHNOIKAMEKRLAEKDEMEQAKRNHLRVVDST 1597
 QY 1382 -----KEEME--FLDELKMASEVEKLFQKA 1405
 DB 1598 QSLDAETRSRNEALRVKRMKGDLNEMEIQLSHANMAEA 1639

RESULT 5

MSB_HUMAN
 ID MSB_HUMAN STANDARD: PRT: 1935 AA.
 AC P12883; Q14904; 516579.
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
 GN MYH7 OR MYHCB.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91065634.
 RA JAEENICKE T., DIEBERICH K.W., HAAS W., SCHLEICH J., LICHTER P.,
 RA PFORDT M., BACH A., VOSBERG H.P.;
 RT "The complete sequence of the human beta-myosin heavy chain gene and
 RT a comparative analysis of its product."
 RL GENOMICS 8:194-206(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90301496.
 RA LIEW C.-C., SOLE M.J., YAMAUCHI-TAKIHARA K., KELLAM B.,
 RA ANDERSON D.H., LIN L., LIEW J.;
 RT "Complete sequence and organization of the human cardiac beta-myosin
 RT heavy chain gene."
 RL NUCLEIC ACIDS RES. 18:3647-3651(1990).
 RN [3]
 RP SEQUENCE OF 1-113 FROM N.A.
 RX MEDLINE: 89264452.
 RA YAMAUCHI-TAKIHARA K., SOLE M.J., LIEW J., ING D., LIEW C.-C.;
 RT "Characterization of human cardiac myosin heavy chain genes."
 RL PROC. NATL. ACADE. SCI. U.S.A. 86:3504-3508(1989).
 RN [4]
 RP ERRATUM.
 RA YAMAUCHI-TAKIHARA K., SOLE M.J., LIEW J., ING D., LIEW C.-C.;
 RL PROC. NATL. ACADE. SCI. U.S.A. 86:7415-7417(1989).
 RN [5]
 RP SEQUENCE OF 1310-1935 FROM N.A.

- RA MEDLINE: 86176778.
 RX SAEZ L., LEINWAND L.A.;
 RT "Characterization of diverse forms of myosin heavy chain expressed in
 RT adult human skeletal muscle.";
 RL NOCTIC ACIDS RES. 14:2951-2969(1986).
 RN [6]
 RP REVISIONS.
 RA LEINWAND L.A.;
 RL SUBMITTED (MAR-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [7]
 RP SEQUENCE OF 1410-1935 FROM N.A.
 RX MEDLINE: 86299163.
 RA KURABAYASHI M., TSUCHIMOTOCHI H., KOMURO I., TAKAKU F., YAZAKI Y.;
 RT "Molecular cloning and characterization of human cardiac alpha-
 RT beta-form myosin heavy chain complementary DNA clones. Regulation of
 RT expression during development and pressure overload in human
 RT atrium.";
 RL J. CLIN. INVEST. 82:524-531(1988).
 RN [8]
 RP SEQUENCE OF 785-1935 FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RX MEDLINE: 90235862.
 RA BOBER E., BUCHBERGER-SEIDL A., BRAUN T., SINGH S., GOEDDE H.W.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL EUR. J. BIOCHEM. 189:55-65(1990).
 RN [9]
 RP SEQUENCE OF 1393-1935 FROM N.A.
 RX MEDLINE: 87192738.
 RA JANDRESKI M.A., LIEW C.-C.;
 RT "Construction of a human ventricular cDNA library and
 RT characterization of a beta myosin heavy chain cDNA clone.";
 RL HUM. GENET. 76:47-53(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE: 96039076.
 RA ARAI S., MATSUOKA R., HIRAYAMA K., SUKURAI H., TAMURA M., OZAWA T.;
 RA KIMURA M., INAMURA S.-I., FURUTANI Y., JOH-O K., KAWANA M., TAKAO A.;
 RA HOSODA S., KOMMA K.;
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
 RT hypertrophic cardiomyopathy.";
 RL AM. J. MED. GENET. 58:267-276(1995).
 RN [11]
 RP VARIANTS HCM GLU-256 AND ARG-741.
 RX MEDLINE: 93248216.
 RA FANANAPAZIR L., DALAKAS M.C., CYRAN F., COHN G., EPSTEIN N.D.;
 RT "Missense mutations in the beta-myosin heavy-chain gene cause central
 RT core disease in hypertrophic cardiomyopathy.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 90:3993-3997(1993).
 RN [12]
 RP VARIANT HCM GLN-403.
 RX MEDLINE: 90367131.
 RA GEISTERFER-LOWRANCE A.A.T., KASS S., TANIGAWA G., VOSBERG H.-P.;
 RA MCKENNA W., SEIDMAN C.E., SEIDMAN J.G.;
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta
 RT cardiac myosin heavy chain gene missense mutation.";
 RL CELL. 62:999-1006(1990).
 RN [13]
 RP VARIANTS HCM GLN-249; GLN-403; CYS-453 AND MET-606.
 RX MEDLINE: 92204193.
 RA WATKINS H., ROSENZWEIG A., HWANG D.S., LEVI T., MCKENNA W.,
 RA SEIDMAN C.E., SEIDMAN J.G.;
 RT "Characteristics and prognostic implications of myosin missense
 RT mutations in familial hypertrophic cardiomyopathy.";
 RL NEM ENGL. J. MED. 326:1108-1114(1992).
 RN [14]
 RP VARIANTS HCM GLN-403; CYS-453; ARG-584 AND MET-606.
 RX MEDLINE: 94070863.
 RA WATKINS H., THIERFELDER L., ANAN R., JARCHO J., MATSUMORI A.,
 RA MCKENNA W., SEIDMAN J.G., SEIDMAN C.E.;
 RT "Independent origin of identical beta cardiac myosin heavy-chain
 RT mutations in hypertrophic cardiomyopathy.";
 RL AM. J. HUM. GENET. 53:1180-1185(1993).
 RN [15]
 RP VARIANTS HCM GLN-403 AND VAL-908.
 RX MEDLINE: 92346810.
 RA EPSTEIN N.D., COHN G.M., CYRAN F., FANANAPAZIR L.;
 RT "Differences in clinical expression of hypertrophic cardiomyopathy
 RT associated with two distinct mutations in the beta-myosin heavy chain
 RT gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";
 RL CIRCULATION 86:345-352(1992).
 RN [16]
 RP VARIANTS HCM LEU-403 AND TRP-403.
 RX MEDLINE: 94075629.
 RA DAUSSE E., KOMADA M., FETLER L., DUBOURG O., DUFOUR C., CARRIER L.,
 RA WISNIEWSKY C., BERCOVICI J., HENGSTENBERG C., AL-MAHDAMI S.;
 RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and
 RT identification of a hot spot for mutations in the beta-myosin heavy
 RT chain gene.";
 RL J. CLIN. INVEST. 92:2807-2813(1993).
 RN [17]
 RP VARIANT HCM TRP-403.
 RX MEDLINE: 94093568.
 RA MOOLMAN J.C., BRINK P.A., CORFIELD V.A.;
 RT "Identification of a new missense mutation at Arg403, a Cpg mutation
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in
 RT hypertrophic cardiomyopathy.";
 RL HUM. MOL. GENET. 2:1731-1732(1993).
 RN [18]
 RP VARIANT HCM ASN-615.
 RX MEDLINE: 93038688.
 RA NISHI H., KIMURA A., HARADA H., TOSHIMA H., SASAZUKI T.;
 RT "Novel missense mutation in cardiac beta myosin heavy chain gene
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";
 RL BIOCHEM. BIOPHYS. RES. COMMON. 188:379-387(1992).
 RN [19]
 RP VARIANT HCM GLY-778.
 RX MEDLINE: 93343938.
 RA HARADA H., KIMURA A., NISHI H., SASAZUKI T., TOSHIMA H.;
 RT "A missense mutation of cardiac beta-myosin heavy chain gene linked
 RT to familial hypertrophic cardiomyopathy in affected Japanese
 RT families.";
 RL BIOCHEM. BIOPHYS. RES. COMMON. 194:791-798(1993).
 RN [20]
 RP VARIANT HCM VAL-908.
 RX MEDLINE: 93168485.
 RA AL-MAHDAMI S., CHAMBERLAIN S., CLELAND J., NIOYANNOPOULOS P.,
 RA GULLIGAN D., FRENCH J., CHOUDHURY L., WILLIAMSON R., OAKLEY C.;
 RT "Identification of a mutation in the beta cardiac myosin heavy chain
 RT gene in a family with hypertrophic cardiomyopathy.";
 RL BR. HEART J. 69:136-141(1993).
 RN [21]
 RP VARIANT HCM TRP-719.
 RX MEDLINE: 95179132.
 RA GREVE G., BACHINSKI L., FRIEDMAN D.L., CZERNUZEWICZ G., ANAN R.,
 RA TOWBIN J., SEIDMAN C.E., ROBERTS R.;
 RT "Isolation of a de novo mutant myocardial beta MHC protein in a
 RT pedigree with hypertrophic cardiomyopathy.";
 RL HUM. MOL. GENET. 3:2073-2075(1994).
 RN [22]
 RP VARIANTS HCM CYS-513; ARG-716, AND TRP-719.
 RX MEDLINE: 9411036.
 RA ANAN R., GREVE G., THIERFELDER L., WATKINS H., MCKENNA W., SOLOMON S.,
 RA VECCHIO C., SHONO H., NAKAO S., TANAKA H., MARES A. JR., TOWBIN J.A.,
 RA SPIRITO P., ROBERTS R., SEIDMAN J.G., SEIDMAN C.E.;
 RT "Prognostic implications of novel beta cardiac myosin heavy chain gene
 RT mutations that cause familial hypertrophic cardiomyopathy.";
 RL J. CLIN. INVEST. 93:280-285(1994).
 RN [23]
 RP VARIANT HCM THR-797.
 RX MEDLINE: 96047159.
 RA MOOLMAN J.C., BRINK P.A., CORFIELD V.A.;
 RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
 RT myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL HUM. MUTAT. 6:197-198(1995).

RN [24]
 RP VARIANT HCM CYS-453.
 RA MEDLINE: 96209901.
 RA KO Y.-L., CHEN J.-J., TANG T.-K., CHENG J.-J., LIN S.-Y., LIU Y.-C.,
 RA KUAN P., WU C.-W., LIEN W.-P., LIEN C.-C.,
 RT "Malignant familial hypertrophic cardiomyopathy in a family with a
 RT 453Arg->Cys mutation in the beta-myosin heavy chain gene:
 RT coexistence of sudden death and end-stage heart failure."
 RL HUM. GENET. 97:585-590(1996).
 RN [25]
 RP VARIANTS HCM THR-349 AND TRP-719.
 RA MEDLINE: 98204402.
 RA JESCHKE B., UHL K., WEIST B., SCHRODER D., MEITINGER T.,
 RA DOHLEMAN C., VOSBERG H.-P.,
 RT "A high risk phenotype of hypertrophic cardiomyopathy associated with
 RT a compound genotype of two mutated beta-myosin heavy chain genes."
 RT
 Query Match 1.9%; Score 146; DB 1; Length 1935;
 Best Local Similarity 16.9%; Pred. No. 1.2;
 Matches 238; Conservative 207; Mismatches 529; Indels 434; Gaps 51;
 QY 203 DGEPPVPRNAYVO-----NFMKDFSGKWTSGTSGNPDPAFCQLHEF 248
 DB 461 DAGEFIDFNSFEOCLINFNEKLOFFNHMFVLEOEYKKEGIEWTFIDFGMDLQAC 520
 QY 249 HMEKDKLVG-----NLTRIKTLGDFFTSAVQTFVODPDLPGALYNHD 293
 DB 521 IDLIEKPMGMSILBEECMFPRKATDMTFKALFD-----NHL 557
 QY 294 NEFLYODDWYLLSIOENKPD-DYIEVYGRND-----AMDYGGSVYTRSPLPST 348
 DB 558 GKSANFOK-----PRNIKGPPEAFSLHYAGIVYNIIGW-----LOKNKDPINTEV 605
 QY 349 INNLKAA-KSYGRPFNNITDNGSGPPLVERLEKTAEGEKLLKEAVEIEEEK 407
 DB 606 VELYOSSSKLSTLFAN-----AGADAP-TEKGKAKKSSFOYVSALH-RENLNK 657
 QY 408 EKEVAVDEMTLFOELLEFKELQDEENFVELSKEKE-----INLELOMEAT- 458
 DB 658 LMTNLRST-----HPHFVRCILIPNETKSGVMDNPLVMHQLCNGVL 699
 QY 458 EYKELFGRLP-----IKRLNALAPHSNPLANHETIKYVSGSKLPGHRESWG 506
 DB 700 EGIRICRKGFPRILYGFRORYRILINPAIPGOFIDSRKGAERLLSLDIDHNOYKFG 759
 QY 507 WEDYGSIVYAKICSRRIIPRFKSPRICCGDLSRGQLFSGHNLSPAS-----INON 563
 DB 760 HTRVFFKAGLGLLEBMRDERLSRIITRI--QAOSRGVLAEMEYKLLERRDSLVIQWN 817
 QY 564 VPKNSGCKFPKPDVALMVEKMGOPAKTAIVAIFLSVASKADAVDALCTCLKEC-- 622
 DB 818 IR-----AFMGYKNN-PMMKLYFKIKPLKSKERKEMASKMEETRLKALE 864
 QY 622 -----RLBLAKISNPACANVACLOTNNRPDETCQIKCGDLFENVY-----VDFENE 671
 DB 865 KSEARKKELEKMWVSLQEKNDLQLOVAEOONLADAEBCDQLIKNKIQLEAKYEMNE 924
 QY 672 -----CAVSRK---KCVPRKSDVGPVDPSPVLVOKDKMDFSKMFTITGL 716
 DB 925 RLEDEEMNAELTAKRKLEDECSSELKRDIDLETLAKVERK----- 969
 QY 717 NPTFADFCDLHEFTEENKVLGNLSWRIRTPDGGFTSAVOKVODPKYPGILYNDN 776
 DB 969 -----HATENK-VKNLEEMAGLD-----ELIATLKEKKAQAEHQAL 1007
 QY 777 EYLLYDDWYLLSSKVENSPEDYIEVYGRNDAMDYGGSVLYTRSA----- 825
 DB 1008 DDLOAEEDKYNTLLTKAKVLEQOV-----DDLEGSLEOEKVMLEPAKRKLE 1056
 QY 825 ---VLPESIIPELOTAQKVG-----RDF-----NTFIDTMCGE-----PLVE 863
 DB 1057 GDILTOESIMLENDKQDLERLKKRDFELNALNARIEDQALSGOLOKLKELOARIE 1116

QY 864 RLEKVEEGEFT-----IKEVEIEEVEK----- 890
 DB 1117 ELEBEL-ESER/ARAKVELRSDLSRELEISRLLEAGATSVQIENKKREAEFOKMR 1175
 QY 890 -----VRQEVTLFSLFEGRELORDEENLRLSKEMVLDGKME 933
 DB 1176 RDLLEATLOHEATAALRRKHADSVALEQIDNLORVOKLEKESSEKLEDD---V 1231
 QY 934 ATEVEKLFGRALPIRLMAVATCFSPCHDRIRFFSSDDGIGRLGITRRINGTEFLKI 993
 DB 1232 TSNMEQIRAKKALEK-----CRT--LEDQNNERS-----KAEIORSVNDL----- 1274
 QY 994 LPPIQADLRTTGGSSRPL---SAFRSGFSKGFIDVPLPSKNEIKELTAPLLKLVG 1050
 DB 1274 --TSORAKIQTFENGELSRLQDEKALISQITRG-----KITYOQLEDLKLPLEEVRKAK 1326
 QY 1051 LACAFILVPSADAVDALTKACCLLKGRILEAKCIANPACANVACLOTNNRPDETCQ 1110
 DB 1327 NALAHALQSAHDDCLLREQYEETAKAELOVLSK--ANSEVAQWT----- 1374
 QY 1111 IKGDLFENSVDDEFNECAVSRKCVPRKSDGEPAPDPVLYQNFNISDFNGWYTS 1170
 DB 1374 -----KRETAIORTEBLEAKKLAORLOEAE-----AVE 1405
 QY 1171 GLNPTFADFCDLHEFTEENKVLGNLSWRIRTPDGGFTSAVOKVODPKYPGILYN 1230
 DB 1406 AVNAKCSLEKTKHRLONEIEDLMVD-----VERSNAAALADKKQ---RN 1448
 QY 1231 HDNEYLHYODDWYLLSIOENKPDYIEVYGRNDAMDYGGSVYTRSPLPST 1290
 DB 1449 FDKILAEWKQKYEESSELSQKE-----ARSLTELFKRAYEES 1491
 QY 1291 LE-----KAASIGRPFNFRTDNTGPEPALVERIEKTYVEGERTIYKEVEIEEY 1344
 DB 1492 LEHLETKRENKNDIEISDITQOLSSGKTHELKVRKQLEAKMELOALFEAEASL 1551
 QY 1345 EKEVEKGRTEMTLFO-----RLAEGENELKODEENFVELS----- 1382
 DB 1552 EHEGKILRAQLFENQJKAIEKRLAEKDEMBQAKRNLHLYVDSIQTSLEAFTRSRNEA 1611
 QY 1382 ---KEEME-FLDEIKMEASVEKLFGKA 1405
 DB 1612 LRVKMKMGDLNEMEIOLSHANMAEA 1639
 RESULT 6
 NM1_YEAST STANDARD; PRT; 2748 AA.
 AC 000402;
 DT 01-DEC-1992 (REL. 24, CREATED)
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE NUCLEAR MIGRATION PROTEIN NM1.
 GN NM1 OR YDR150W.
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMETES; SACCCHAROMYCETALES;
 CC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 28383 / FL100;
 RX MEDLINE: 92079907.
 RA KORMANEC J., SCHAEFF-GERSTENSCHLAGER I., ZIMMERMANN F.K.,
 RA PERECKO D., KUENZEL H.,
 RT "Nuclear migration in *Saccharomyces cerevisiae* is controlled by the
 RT highly repetitive 313 kDa NM1 protein."
 RL MOL. GEN. GENET. 230:277-287(1991).
 CC -1- FUNCTION: CONTROLS NUCLEAR MIGRATION. NM1 SPECIFICALLY CONTROLS
 CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
 CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
 CC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
 CC ENVELOPE.
 CC -1- ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT CONSENSUS
 CC (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH FLANKING

CC DOMAINS OF THE TANDEM REPEATS.
CC - SIMILARITY: CONTAINS 1 PH DOMAIN.
CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X61236; G4072; -
DR PIR: S19052; S19052.
DR SGD: L0001287; NDM1.
DR PROSITE: PSS0003; PH_DOMAIN; 1.
DR PFAM: PF00169; PH; 1.
KW REPEAT.
FT DOMAIN 593 1384 12.5 X TANDEM REPEATS.
FT REPEAT 593 656 1.
FT REPEAT 657 727 2.
FT REPEAT 728 798 3.
FT REPEAT 799 862 4.
FT REPEAT 863 926 5.
FT REPEAT 927 990 6.
FT REPEAT 991 1054 7.
FT REPEAT 1055 1118 8.
FT REPEAT 1119 1182 9.
FT REPEAT 1183 1246 10.
FT REPEAT 1247 1310 11.
FT REPEAT 1311 1374 12.
FT REPEAT 1375 1384 13 (INCOMPLETE).
FT DOMAIN 2573 2683 PH.
SQ SEQUENCE 2748 AA; 313202 MW; 3085662C CRC32;

Query Match 1.9%; Score 143; DB 1; Length 2748;
Best Local Similarity 17.9%; Pred. No. 2.9;
Matches 283; Conservative 231; Mismatches 582; Indels 486; Gaps 72;

QY 62 SSSSDSHCKDQKQSCISDITSEFIDRFIDKRGMTILEKQNRQF-----IQALIVLC 115
DB 748 SAEDLVCKENPDV-----EFLKEKSAKLGHTVVSSEYSELQKYSLEKEVEQPS 800
QY 116 TEYIVPRVDADLKTACCLKECRILEAKCIANPSC-----AANVACLOTCNNPDET 169
DB 801 LALIVEHAKATDH-----HLSDSAIELVCKENPDMEFLKEKSAKLGHTVVSNEAYSEL 856
QY 170 EQGI-----KCGD--LFENSVDQFNECAVSRKKCVPRKSDVGEFPPVDNRNAV 215
DB 857 EKLLEOPSLAYLVEHAKATDHLLSDAYEDLYK-----KENSVDY-EF----- 900
QY 216 VQNFNNKDPGKMYITSGINPTFDADCOLHEFHMDKLVGNLTIRITLDCGFFTRSA 275
DB 900 -----LKEKSAKLGHTVVSNEAYSELEKLEQ-----PSLAYLVEHAKATDHLLSDSA 948
QY 276 VQTFVODPDLPGALYHNDHEFLHYQ-----DDWYILSSOLENKPDDYIFVYRGA 324
DB 949 YE-----DLVKCKENPDMEFLKEKSAKLGHTVVSNEAYSELEKLEQPSLEYVEHAKA 1002
QY 325 RND---AMDYGGGV-----LYTRSPLLPESILFN-----LOKAAKGVGDF-- 364
DB 1003 TNNHLLSDSAIEDLVCKENPDMEFLKEKSAKLGHTVVSNEAYSELEKLEQPSLEYVE 1062
QY 364 -----NNFITDSS-----CGEPPLVERLEKTAEGEKKLKEAV--EIEEVEKE--- 409
DB 1063 HAKATHHLLSDSAIELVCKENPDVFLKEKSAKLGHTVVSNEAYSELEKLEQPSLE 1122
QY 409 --VEKVRDEMTLFFORLGEFKELODEENFVRELSEKEKEILNE--LDMEA-TEVEKLF 463
DB 1123 YLVEHAKATNNHLLSD--SAIELVCKENPDVFLKEKSAKLGHTVVSNEAYSELEK-- 1179
QY 464 GRALPILRKLMALAPHSNLANHETIKYVGSKLPGHKRFSGWEDYEGSIVAVK----- 519

DB 1179 -----KLEO---PSLAYLVEH-----AKATDHLLS-----DSAYEDLVCKENPDV 1217
QY 519 -----ICSSRRILPRYPRKSPRICCGDSHGL-QLPSHGHNLSPAHSINOV 564
DB 1218 EFLKEKSAKLGHTVVSNEAYSELEK-----LEOPSLAYLVEHAK--ATDHLLSDSA 1268
QY 565 PKNSGCKFPDVALMWEKMGQPAKTAIYAIFLLVASKAD-----AVDALKTCTCL 617
DB 1269 YEDLVCKENPDMEFLK-EKSAKLGHTVVSNEAYSELEKLEQPSLEYVEHAKATNNHLL 1327
QY 618 LKEORLE-LAKCISNPAC-----AANVACLOTCNNPDETEQICGDFENSVDEN 670
DB 1328 LSDAYEDLVCKENPDMEFLKEKSAKLGHTVVSNEAYSELEKLEQPSL----- 1378
QY 671 BCAYSRKKCVPRK--SDVGF-----PVPPSYLVQFKMDKDSGKWFITRGINPTFDAF 723
DB 1378 EYLVKHAEQIQSKLISLSDNTLANPSMEDMAKQLKLEQIYSNDEYI--ALKNTMEKP 1435
QY 724 DCQ-----LHEFH-----TEENKLVGNISWRIRTPDGGFFTRSAVQKFPVD--PRYPGILYN 773
DB 1436 DVELLRSLKAGYHLLDTTYNELVSNFN-----SEPLKFLKEKAKSKGYRLI 1482
QY 774 HDNEVYL-----LYODDWTLSK-----VENSPE-----D 798
DB 1483 EPNEYLDNLRIATTPSKSEELDNCKQIGYALDSKEYERLKNLENPSSKFTIENALLD 1542
QY 799 YIFV---YKGRNDAMDYGGSVLYTRSAVLPSIIPELQTAQKVGDFNFPIKTIDNTC 855
DB 1543 LVLDKTEYQAMKD-----NASNKSILPSTKVL-----DEVEN----- 1577
QY 856 GPEPPLVERLEKAYEBSERTTIKEVEEIEEVEKAVKDEVTLESKLPEGKELQRD---- 912
DB 1577 -PAPOLSAEKSSLOK-----RTLSDIENELKAL-----GYVALRKKNLEPN 1616
QY 912 -EENFLRELKSEEM-----DVLDELKMEATEVEKLF-----RALPIR 948
DB 1617 LEKIVYNAKSNVDYLNLCNSKFSVLPLSTEDYDMKREHKILNILLDPSTDLKEKCEY 1676
QY 949 KLMVAATH-----CFTSCHDRIRFFSSDGIIGRLITRKRIINGTFLKILPIQSD 1001
DB 1677 QMLIISHDEEKOEALENQYEFILKEKASALGY-----ELVSVEILDPMQMDSPD 1729
QY 1002 LRTGGRSPLSAFRSGFSGKGIPIVPLPSKNELELAPLLKLVGLACFLVPSA 1061
DB 1730 IDIYQEKAAARN-----EMVLL--RNEEKE---ALQKKIEYPSLTFLEKAA 1770
QY 1062 -----DAVDALKTACCLKECRILEAKCIANPACANVACLOTCNNPDETEQICG 1114
DB 1771 GNNKILVDQIEYDET-----IRKC-----NHPTRHELESCH 1802
QY 1115 DLFENSVDQFNECAVSRKKCVPRK--SDGEFPAPDPSVLYVONFINSDNGKMYITSGIN 1173
DB 1803 HL--NLVLLQNEKSTLREPLKRNVEDLINTLSKINYAIPNTIYQDLIGKYE-----N 1855
QY 1174 PTFDAPFCOLHEFHTEGDNKLVGNISWRIKTLDGFFTRSAVQKFPQDNOGVLY---- 1230
DB 1856 PNFYLDLSLAK-----MDYVAISRQDVELYAKAKEKRPOLDYKLS 1896
QY 1230 -----NHDEYLYQDDWTLSKIEKNKEDYIFVYRGRNDAMD 1270
DB 1897 SEKIDHIVPLSEYNLMTVYRNPSTLYLKEKAVLNNHILIKEDDYKNLL-----AYSE 1950
QY 1271 YGCAVVTTRSSVLPNSIIPLEKAASIGRDESTFRTDNTCGPEPALVERLEKTYEBS 1330
DB 1951 HPTVILHSERHLLINKVLVD-----RNDFAITMSRSIE---KPTIDFLSTKALSMG- 1998
QY 1331 RIIVKEV-----EIEEVEKEVEKVGRTMETLFOGLAGFNELKODEENFVRELSEK 1385
DB 1998 KIIYNSTHTRNKLSEPSSEFLTKAKAEOGLITSEKYSLSLRQIDRPNDLVLEKKA 2057
QY 1386 EFLDEIKMEASEVEKLEGRALP 1407

Db 2058 AIFDSIIVENYQOLVNTSP 2079

```

RESULT 7
MYSB_RAT STANDARD; PRT; 1935 AA.
ID MYSB_RAT
AC P02564;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH7
OS RATTUS NORVEGICUS (RAT).
OC EUDAROTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCURIONATHI; MORIDAE; MORINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA KRAFT R., BRAVO-ZEHNDER M., TAYLOR D., LEINMANN L.A.;
RT "Complete nucleotide sequence of full length cDNA for rat beta
RL NUCLEIC ACIDS RES. 17:7529-7530(1989).
RN [2]
RP DISCUSSION OF SEQUENCE.
RA MCNALLY E.M., KRAFT R., BRAVO-ZEHNDER M., TAYLOR D., LEINMANN L.A.;
RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
RT Comparisons suggest a molecular basis for functional differences."
RL J. MOL. BIOL. 210:665-671(1989).
RN [3]
RP SEQUENCE OF 1524-1935 FROM N.A.
RA MAHDAVI V., PERLASAMY M., NADAL-GINARD B.;
RT "Molecular characterization of two myosin heavy chain genes expressed
RL in the adult heart."
RN NATURE 297:659-664(1982).
RN [4]
RP SEQUENCE OF 1871-1935 FROM N.A.
RA STRAIN-WISTAR; TISSUE=HEART;
RC MEDLINE; 85179510.
RA MAHDAVI V., LOMPRE A.M., CHAMBERS A.P., NADAL-GINARD B.;
RT "Cardiac myosin heavy chain isoform transitions during development
RT and under pathological conditions are regulated at the level of mRNA
RT availability."
RL EUR. HEART J. 5:181-191(1984).
RN [5]
RP FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC DOMAINS OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM)
CC AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE SPLIT FURTHER INTO
CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE MYOSIN, WHILE THE
CC BETA ISOFORM IS A 'SLOW' ATPASE.
CC -1- THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN THE RAT, TWO OF
CC WHICH ARE SPECIFIC FOR ADULT CARDIAC MYOSIN HEAVY CHAINS.
CC -1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
CC CONSERVED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/

```

```

or send an email to license@isb.sib.ch).
CC -----
CC EMBL: X15939; G54657;
CC EMBL: J00752; G205578;
CC EMBL: M32698; G205599;
CC PIR: S06006; S06006.
CC PIR: A02989; A02989.
CC PIR: P00063; myosin_head; 1.
CC PIR: P00612; IQ; 1.
CC HSP: P08799; LMD.
CC MYOSIN: MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING.
CC ATP-BINDING; METHYLATION; ALKYLATION; MULTIGENE FAMILY.
CC KX MYOSIN: MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING.
CC KW MYOSIN: MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING.
CC FT DOMAIN 1 839
CC FT DOMAIN 840 1935
CC FT DOMAIN 780 802
CC FT DOMAIN 840 1935
CC FT NP_BIND 178 185
CC FT DOMAIN 655 677
CC FT DOMAIN 757 771
CC FT MOD_RES 129 129
CC FT MOD_RES 695 695
CC FT MOD_RES 705 705
CC FT MOD_RES 1529 1531
CC FT CONFLICT 1731 1731
CC FT CONFLICT 1784 1784
CC FT CONFLICT 1851 1851
CC FT CONFLICT 1858 1858
CC FT CONFLICT 1935 NA; 223082 MW: 1D16FF6E CAC32.
CC SEQUENCE 1935 NA; 223082 MW: 1D16FF6E CAC32.

Query Match 1.98; Score 142.5; DB 1; Length 1935;
Best Local Similarity 16.7%; Pred. No. 1.9;
Matches 247; Conservative 218; Mismatches 558; Indels 459; Gaps 56;

166 PDETCQIKRCQJLENSVVDQFNCAVSRKKVPRKSDVEEFPVDRN-----AVV 216
Db 375 PDGTE-----EADKSAVLMGLNSADLLKGLCHPRKAVGNEYTKQAOVAYAGALA 428
OY 217 QNFNKKDFSGKRYITSGNPT-----FDAFDCOLHE---FHMENDKL 255
Db 429 KSVYKRMN--NMVTR--INATLETKOPROYFIVGLDIAEIEIDFNSFQGLCNFTNEXL 485
OY 256 VGNLTWRKIKTLDGFFETRSAYO--TFVQ-----D 282
Db 486 QGFENHMFVLEOEDEYKKEGEMTFIDFGMDLQACIDILEKPMGINSILEECMPKRAID 545
OY 283 POLPEBALY--NHQNEFLHODQDWYLLSSQIENKPPDYIFVYRRGND---AMDYGGASY 337
Db 546 MTFKAKLYDNHGLKSNNOQKPRNIKGOEAB---SLIHAGTVYNIIGW---L 594
OY 338 YTRSPTLPESTIPNLQKAA--KSVGRDFNFTITDNSCGPEPPYERLEKTAEGEKLLIK 396
Db 595 QNKKPPLNETVYVGLQKSLKLSNLFANY-----AGADAP--VDKQKAKKAGSSFFQV 647
OY 397 EAVEIEBEYKVEKEVVDTEMTLRLQLEGKELQODEENFVRLSKENE-----448
Db 648 SALH--RENINK--MTMIRST-----HPEVRCIIPNETKSGVMDNPL 668
OY 448 LNELQMENT--EVLTFGRAP-----IRKLPMALPNSFLANHETIKYVGS 495
Db 689 VMHQLKNCVLEGIKICKRGPNNRLIYDFQRRIRLNPALIPSGQITDRKAGKELKGS 748
OY 496 KLPGHKRFSGMEDYFGSIYVAKICSSRRIPRYFRKSPRILCCGDSGLQIFSHGHNLS 555
Db 749 LDIDNHQYFQHTVFPAAGLLGLEEKRDRLSRIIRI--QASRGVLSRMEFFKLL 806
OY 556 PAHS---INQNPKNKSGCKEPKVDALVMWEKGFATVAFILISVASKADAV---609
Db 807 RRDSDLIIDWNR-----AFMGVKNWPMKLYEIKPLLSAETKEMANKE 854
OY 609 -----DALKTCTCLKECRLELACISNPACANVACLOTNNRPDETQCIKCGDLFE 662
Db 855 EFORVYDALEKSEARKKELEEMNVSLLQE-----KNDLQLOVQAQDNLAAEERCGDLIK 910

```

```

OY 663 NSV-----VDFENE-----CAVSRK---KCVPRKSDVDFVPPSVLVQFED 702
DB 911 NKIOLEAKKEMTERLEDEEENKALTTAKRKLEDCSLKRDIDDELTTAKVKEK-- 969
OY 703 MDFSCKMFTITGLNFTEDAFDCOLHEFTENKLVGNLSMRIRTPDGFFTRSAVQFV 762
DB 969 -----HATENK-VKNLTEMALD-----ELIVKLT 993
OY 763 ODPKYPGILYNHNDNEYLLODDMYILSSKVENSPEDYIFVYKGRNDAMGYGGSVLYTR 822
DB 994 KKKKALQEAHQOALDLOAEEDKRVNLTAKVKEQOV-----DDLEGSIDDK 1042
OY 823 SA-----VLESILPELOTAOKVG-----RDF-----NTIKINDICGP 857
DB 1043 KVRMDLERAKRKLEGDINKTOSINDLENDKQOLDERKCKDFELNANARIEDQALGS 1102
OY 856 E-----PPLVERLEKKEVEGERT-----IKEVEETEVEVER----- 890
DB 1103 QLOKKIKELQARIELEBEL- EAERTARAKVEKLRSDLSRELEISERLLEAGATSVOI 1161
OY 890 -----VRDKEVTLFSKLFEGFKLODEENFIREL 919
DB 1162 EBNKREAFQKMRDLEBATOATATAALKKHADSVAGIDNLOVKOKLEK 1221
OY 920 SKEMDVLDGLKMEATEVEKLFGRALPIRKLAVAHCTSPCHDIRIFFSSDGIKRLG 979
DB 1222 SEFKLELD-----VTSNMEQITKAKANLEK-----CRT--LEDQNHHS-----KAE 1264
OY 980 ITRKINFTFLKILPIQASADLRITGGRSRPL-----SAFSGSGSKGIFDIPLPSKNEI 1036
DB 1265 ETORSAVNL-----TRORAKLOENELSRQDEKALISQLTRG-----KULTYQOL 1312
OY 1037 KELAPILKLVGLACAFLLVPSADAVDAKTCACILKGRILACIANPACAAVAC 1096
DB 1313 EDLKRQLEEEVYKAKNALAHALQSAHODDLREQEETEETKALQVLSK--ANSEVAQ 1370
OY 1097 LOTCNRRDETECOIKCGDLFENSVDVEFNCAVSRKCVPRKSDLEFPAPPSVLYQN 1156
DB 1371 WRT-----KYETDAIQRTBELLEAKKKLQRLQDAE----- 1403
OY 1157 FNISDENGKWTITGLNFTEDAFDCOLHEFTEDGNKLVGINSRITLDSGFFTRSAVQ 1216
DB 1403 -----AFAVNAKCSLEKTKHRLQNETEDLVD-----VERSNA 1438
OY 1217 KFOVDPNPOGVLYNHNDNEYLAHODWYILSSKLENKPEDYIFVYGRNDAMGYGAVY 1276
DB 1439 AAALDKQ-----RNFDKILVEMKQKYEESQSELESQK-----ARSL 1477
OY 1277 YTRSSVLPSNIIPELE-----KAASIGRDFSTFTRDNTCGEPALVERIEKTYEGE 1330
DB 1478 STELFKLNAVEESLEHLETFKRENKNOLEIEISDLTEBOLSTGSIHELEKIRKQLEAEK 1537
OY 1331 RIIVEVEEIEEVEKEVEKVEVGRTEMFLFO-----RLAEGFENKODEENVRELS-- 1382
DB 1538 LELOSALEEAASLEHEGKILRAQLEFNOIKAEIEKLAKEDEEMQAKRNLRAVDSI 1597
OY 1382 -----KEDME-FLDEIKMEASEVEKIFKRA 1405
DB 1598 QTSIDAEATRSRNEALRYKKRMEGLNEMEIQLSHANMAEA 1639

```

```

OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-6288C;
RX MEDLINE: 93247549.
RA KOELLING R., NGUYEN T., CHEN E.Y., BOSTERIN D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RM MOL. GENET. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94205265.
RA BOU G., ESTEBAN P.F., BALABON V., GONZALEZ G.A., CANTALEJO J.G.,
RA REMACHA M., JIMENEZ A., DEL REY F., BALLESTA J.P.G., REVUELTA J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and Mpl1 genes and three
RT new open reading frames.";
RL YEAST 9:1349-1354(1993).
CC - FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC - SIMILARITY: SOME TO THE TPR ONCOGENE.
CC - CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: L01992; G111959; -
DR EMBL: X73541; G450554; -
DR EMBL: Z28320; G466587; -
DR PIR: S38173; S38173.
DR SGD: L0001122; MPL1.
DR MYOSIN; HEPTAD REPEAT PATTERN; COILED COIL; DNA REPAIR.
FT DOMAIN 69 -- 487
FT DOMAIN 531 1678
FT DOMAIN 1834 1866
FT CONFLICT 301 301 R -> A (IN REF. 1);
SQ SEQUENCE 1875 AA; 218455 MW; 8B01FDD0 CRC32;

```

Query Match 1.88; Score 137.5; DB 1; Length 1875;
Best Local Similarity 17.58; Pred. No. 3.6;
Matches 252; Conservative 240; Mismatches 525; Indels 421; Gaps 67;

```

OY 171 QOIKCGDLFENSVD-----QFNEC-----AVSRKCVPRKSDVEFP 208
DB 30 CSLEGVKSFDDGVYKHLNDKLOFNELKSENLKATVSPFDELKASSLKIDGKTEM-ENV 88
OY 209 VPDRAV-----VONFMKDFSGKWTITGLNP-----TFDAF 241
DB 89 IRENDKIKERNDFVYKFESEVENKMKLSSELEFVKRLDLDEKKETQSONOQRTKIL 148
OY 242 DCOLHEFM-----END-----KLVGNTLWTRKTLDDGEFT-----BSAVGTFOVDDPLG 287
DB 149 DERLKEITLAVAVENNRNSSECKKILRTIMDIETKQOQYITINDLSKRIELEKRTQELTL- 207
OY 288 ALYNHNDNEFLHYODWYILSSQLENKRPDYIFVYGRNDAMGYGGSVYTRSPILPES 347
DB 207 -----LOSNDW--LEKELRSKNEQTL--SYRQKTQ-----K 234
OY 348 IIPNLOKAASVGRDF-----NFTITDNCGEPPLVERLEKTAEGEKLKAEVEIE 402
DB 235 VILDIRNELNLRNDFOMERTINDVLKQKN-----NELSKSLQ--EXLT--ETKGLS 282
OY 403 EBEVEKEKYADTEMTLFFQ---LLEGFKELQODEENFVEL-----SKEEK 446
DB 283 DLSNKEQEF-SAEMSKQRLVLDLLESQLNAVKEELNSITELTAKAVIADSKKQTPENE 341
OY 447 EILNELOMEATEVEKLFGRALPIRKLR-MALADPSNLANHETIKYVSGSKLPGRHFRFSW 505

```

```

Db 342 DLKELQTLKELKAOCEKELRLSSITDEADENENLSAKSSDFELFKOLIKERRKE 401
QY 506 GWEDYFGSIYVAKICSSRIPIYFKPSRICGLDRLQJLFSHGKHLSPAHISINQNP 565
Db 402 HLONOJEIFIVLEHKKYPLINSFKERTDMLLENLNNALLLEHTSNEKNAVKELN--A 458
QY 566 KGNSSGCKFPKDPALWMEKW-----GQFAKTAIVAFIILSVASKA 605
Db 459 KNOKLVEECENDQTLKQDLDCRQIYILITNSVNSDKPLKEELQ--FINQIMQED 516
QY 606 DAV---DALKTCTCLLKECR--LELACISNPACANAVACLOTCNNRPDETCOIKGC- 659
Db 517 DSRTESDSOKVYTERLVKFNILQLOE-----KNAELIKVARNLADLSEKREKSK 568
QY 659 ---DLFENVSDVEFNCAVSRKKCVPRKSDVGD-----FPVPPSVLVQKFD 702
Db 569 QSLQKIESEVNEAKPAITTLKS--EKMDLESRIEELQKLEBELKTSVNPEDASYSMT 625
QY 703 MKDFSGKWFITGLNPTDFADFCOLHEFTEENKLVGLSWIRTPDGGFTFSRNVQFV 762
Db 626 IKQ-----LTFETKRDLESQVODLQTRISQIT--RESTEN-----MSLKNKEI 665
QY 763 QDPKYGILYNHDNEY-----LIYQDWYILS---SKVENSPEDIYFVYIKG 806
Db 666 QD-----LYDSKSDISIKLGRKSSRIIAEERFKLSTLIDLTKAENDQLRKREDYLQN 719
QY 807 -----RNDAMDYGG-----SVLYTSAVLPEESILPELOTAOKGGRDFNFPIKTD- 853
Db 720 TILKQDSKTHELNEYVSCSKSLIVETELLNKE-----QKLVEHLKWLKEL 770
QY 853 NTGCPK-----PLVERLEKKEVEGERTIIEVEEIEEVEKVRKDEVTLLFSKL 901
Db 771 NKLSPEKDSIRIYVQLOTLQKREDBLLETRKSCQKIDELDALSELK-KET---SQX 826
QY 902 FEGFKELQDE-----ENFLBELSKEMDVLDGLKMEATEVEKLFGRALPIRLMLAVA 954
Db 827 DHKIKOLEENNNSNIEMYQKIKALKKDYESVITVSQKQDIEKLOKVKSLKEIE- 885
QY 955 THCFTSCHRRIRFFS-----SDDGIGRLGITRKRIRNGTEFLKILPIPSADLRTFG 1006
Db 885 -----EDKIRLHYNVMDETINDSL-RKELERSKINLNDANSQIK--EYKLIYET- 933
QY 1007 GRSSRPISARSGFSKGIPIVPLPSKNEIKELTAPLLKLVGLVACAFIYPASDAVDA 1066
Db 933 --TSQSIQOTNSKLIDESKFDT-----NQIKNLT-----DEKTS 964
QY 1067 LKTCACILKGCRIELAKCIANPACANAVACLOTCNNRPDETCOIKGDLFEN-----SV 1121
Db 965 LEDKISILKQOMFL-----NNEDLQKKGMEKEKADFKKRISILQNNKKEVEAV 1014
QY 1122 VDENECAVSRKKCVPRKSDLGFEPPADPSVIVQNNISFNKQWITSGMLPTDAPDC 1181
Db 1015 KSEY-----ESKLSKIQNL-----DOOTIYANTANQNNYDOELQKHADVSTTSELRE 1062
QY 1182 QLHEFHTEGDKNLVGNISWRIKTLDGSGFTFSRNVQKFPDPNPQVLYVHNDDEYLYQDD 1241
Db 1063 QLHLY-----KGVKQTLN---LSRDQLENALKKEMKS--WSSQKESLLEQJD 1104
QY 1242 WYILSKIEKPE-----DIIEVYIRGRNDAMDYGCG-----AVYIRSSVLPNSIIP 1289
Db 1105 --LNSRIEDLSSONKLLYDOIQIYTAADKEVNNSNGPCLNNILTLTRERDIDLTKYT 1162
QY 1290 ELEKAASIGRDFSTF-----IRFDNTCGPEPALVERIEKTEVEGERIILYKEVEIE 1341
Db 1163 VAERPAKTLRQKISLMDVELODANTKLQNS-----RYEK---ENHSSITQOHDJIM 1210
QY 1342 EBEVEKEVKGRTENTLFLORLAEGFNEIKODEENFVELSKEMEFLDEIKMEASVE 1399
Db 1211 EKL-NQNLWLRRESNITL-----RNEL-ENNNKKKELQSE---LDKLOVAVIE 1255

```

RESULT 9
DMD_CHICK

```

ID DMD_CHICK STANDARD; PRT: 3660 AA.
AC P11533;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE DYSTROPHIN.
GN DMD.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89098331.
RA LEMAIRE C., HEILIG R., MANDEL J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA.";
RL NUCLEIC ACIDS RES. 16:11815-11815(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE-MUSCLE.
RA LEMAIRE C., HEILIG R., MANDEL J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal
coding and 3' untranslated regions between man and chicken.";
RL EMBO J. 7:4157-4162(1988).
CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
CC PLASMA MEMBRANE.
CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FILBRIN,
CC ABP-120, ABP-180, OR BETA-FODRIN).
CC -1- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X13369; G63370; -.
CC PIR: S02041; S02041.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS00020; WW_DOMAIN_2; 1.
DR PFAM: PF00307; actinln-binding; 1.
DR PFAM: PF00397; ww_rps5_mwp; 1.
DR PFAM: PF00435; spectrin; 22.
DR PFAM: PF00569; z1; 1.
DR HSSP: O01082; 1A42.
DR STRUCTURAL PROTEIN; ACTIN-BINDING; CALCIUM-BINDING; CYTOSKELETON;
KW REPEAT.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 300 3000 26 SPECTRIN-LIKE REPEATS.
FT DOMAIN 3052 3085 WW DOMAIN.
FT DOMAIN 3086 3357 CYS-RICH.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 O -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; AF61A205 CRC32;

Query Match 1.8%; Score 135; DB 1; Length 3660;
Best Local Similarity 17.6%; Pred. No. 12;
Matches 245; Conservative 196; Mismatches 482; Indels 472; Gaps 64;

QY 71 KDSQISITSTSEDEIORDLAKGM--TII-LEKQROF-----IQAIAYVC 115
Db 1923 KKKEDLNAVNPQAEERLSKQDAKAVETIYVLSKRWRFESKFAQFRRLNQAQIQTVED 1982
QY 116 TFEVIVPRVAVDALKTCACILKREIELAKCIANSCANAVACLOTCNNRPDETCOIKC 175
Db 1983 TTFVMTESMTVEVTTYPSTYLAET-IQLQALSEVERLNSVLDQAKD-----C 2030

```


QY 1165 KWTITGLNPTDPAFCOLHEFHTEGDNKLVGNISRIKTLDSGFTFSANQFVQDPNQ 1224
 DB 1413 K---CASLKTORTLONEDMLDVE-----RSNAACALDKQ 1449
 QY 1225 PGVLNHDNEYLYHODDWTLLSKLENKPEDEYFYVYRGNDAMDGYGAAVYVTSYLP 1284
 DB 1450 ---RNFEDVLSEWKKQKYEETQALEASQKE-----SRSLSTELFKVKNVYEESSL--- 1496
 QY 1285 NSIIPLEKAAASIGRDEFTFRTDNTCGEPALVREIKETVEEGEERITVKEVEIEEEV 1344
 DB 1496 -DQLETLRRNNKMLQOEISDLTEQIAEGSQIHELEKIKKQVEQKCEIQALAEAEASL 1554
 QY 1345 EKEVEKVCRTMTLFO-----RLAEGENIKODENVEFLS----- 1382
 DB 1555 EHHEGILRLIOLELNOYKSEVDKIAKDEEIDQLKRHTRVVETMOSTLDAEISRDA 1614
 QY 1382 ---KEEME-FLDEIKKAESEVERLEFGKAL 1406
 DB 1615 LRVKKKMGEDLNEMETIOLNHNRLAESL 1643

RESULT 11
 MSP1_PLAER STANDARD: PRT; 1630 AA.
 ID MSP1_PLAER STANDARD: PRT; 1630 AA.
 AC P04932;
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
 DE (PMMA) (P190).
 GN MSP-1.
 OS PLASMODIUM FALCIPARUM (ISOLATE K1 / THAILAND).
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 86136024.
 RA MACRAY M., GOMAN M., BONE N., HYDE J.E., SCAIFE J., CERTA U.,
 RA STUNNENBERG H., BUJARD H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 RT Plasmodium falciparum merozoites: studies at the genetic level";
 RL EMBO J. 4:3823-3829(1985).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RA PAN W., TOLLE R., BUJARD H.;
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -I- P1M: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KD, 42
 CC KD AND 19 KD ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; X03371; G929798; -
 DR PIR; A25120; SAZOKL.
 DR PIR; P04000; EGF; 1.
 KW MALARIAS; MERZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
 KW TRANSMEMBRANE; GPI-ANCHOR.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1630 MERZOITE SURFACE PROTEIN 1.
 FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
 FT TRASMEM 1614 1630 MEMBRANE ANCHOR.
 FT CARBOHYD 97 97 POTENTIAL.
 FT CARBOHYD 259 259 POTENTIAL.
 FT CARBOHYD 755 755 POTENTIAL.
 FT CARBOHYD 759 759 POTENTIAL.
 FT CARBOHYD 774 774 POTENTIAL.

FT CARBOHYD 835 835 POTENTIAL.
 FT CARBOHYD 911 911 POTENTIAL.
 FT CARBOHYD 955 955 POTENTIAL.
 FT CARBOHYD 1049 1049 POTENTIAL.
 FT CARBOHYD 1156 1156 POTENTIAL.
 FT CARBOHYD 1165 1165 POTENTIAL.
 FT CARBOHYD 1436 1436 POTENTIAL.
 FT CARBOHYD 1517 1517 POTENTIAL.
 SQ SEQUENCE 1630 AA; 187289 MW; DD2F8628 CRC32;

Query Match 1.8%; Score 134; DB 1; Length 1630;
 Best Local Similarity 18.0%; Pred No. 4.7; Mismatches 473; Indels 444; Gaps 70;
 Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

QY 184 VDFNECAVSRKKCVPRKSDVGEFPVDRNAVQVNMKDFSGKWTITGLNPTDPAFC 243
 DB 240 IENINELIEESKTKIDK-----NKNATKKEEKKLYQAQDLS-----IYNK 281
 QY 244 QLEPHENDKLVGNLTWRKTLIDGFTFSAVQTF--VQDP-----DLPGALYNHDNE 295
 DB 282 QLEBAH---NLISVLEKRIIDLTKNNENIKELDLKINEKNPPANSGNTPMTLLDKNKK 337
 QY 296 FLHYODDWTLLSQIENKPPDY-----IFVYRGNDAMDGYGSAVYTRSPITPESII 349
 DB 338 IEHHEKEIKELATIKFENIDSLFTDLELEYLIREKNRID--ISAKYETKSTEPNE-Y 394
 QY 350 PNIQAKASVGRFNFITDNSCGPEPLVERLEKTAEGEKLLI---KEAVEIEEE 404
 DB 395 PNVGYPLSY-NRINNALNELNSFG--DLINPDTKPSKNITYTDNRKKFIEIKER 450
 QY 405 VEKEVEKVRNTEMTLEFORLLEGEKLEQDEENFVELSEKEEILNEL-----QMET 457
 DB 451 IKIEKKKI-----ESDKSYEDRSKSLNDITYEKELNELIYDSFENNIDLT 498
 QY 458 EVELKFR--ALPIRLMALAPHSNFLANHETIKYVYSGKLPGHKRFSGWEDY-FGSI 514
 DB 499 NFEKMGKRSYVEKL-----THHTFASYSKSNHL-EKILTKLKY--MEDSLANI 549
 QY 515 VVAKTSSRRIPYPRKSPRICGLDSRLQLFSGKHNLSAHSINONVPGNSGC--- 572
 DB 550 VVERKL---KYRK-----NLISKIENEI--ETLVENIKKDEEQLPEK 586
 QY 572 ---KPRDVALWMEKMGQFAKTAIYAILLSASADAVDLKTCCLKRCRE- 625
 DB 587 KITKDNKPKDEKLEV---SDIVKYQOKVILM-----NKIDELKTKQLTLKAVELKH 636
 QY 625 -----LAKCISN-----PACAA-----NVACLOTCNNRPDET 651
 DB 637 NIHPVNSYQENKQEPYVLIVLKEIKDLKVFMPKYESLINEKNINITEGGSNSEST 696
 QY 652 ECGI-----KCGDLFENSVDNEFCASRRKKCYPRKSDVGDFFVPPSVLVQ-- 700
 DB 697 EGBITQATTKPQOAGSALG---DSVOAQOEOKQAO--PPVPVPEKAAQVP 747
 QY 700 -----KRDMDFGSKWFTRGMLNPTDPAFCOLHEFHTEENKLVGLNLSRI 745
 DB 748 TPAPAPNNKTEVSKIDYLE-----KIVEFL-----NTSYIC 779
 QY 746 RPPDGFETFSAVQKFFVQPKYPGILYNHD--NEVLYQDDWTLLSKRYEN-----SPED 798
 DB 780 H-----KY--LLVSHSTNNKILKQ--YKITEBESKSCDPLD 815
 QY 799 YIF-----VYKKGNDAMDGYGSAV-----LYTRSAVL-----PESTIPELQITAAQ 839
 DB 816 LFNFIQNNIPVKT---SMFDSLNSLSQLEMEIYEKEWNCVNLKLDNDKTKNLLEBAK 871
 QY 840 KYGRDNTFKITDNTGPP-PLVERLEKKEVEGERT-----IKIVEEII----- 884
 DB 872 KV---STVKTLSSSSMPLSLITPDKREVSANDTISHTLNNLSKLFENILSLGNK 927
 QY 884 -----EEVEKVRDEKVTLSKLFEGEKLQDEENFVELSKEMDVLGDK 931


```

Cc 928 NIYQELIGKSSSENYEKILKDSOTFYNESFTNFVKSADINSLANDESKRK-----K 980
Qy 932 MEATEVEKLFGRALPIRKLMAVAHHCFTSPCHDRIREFSSDDGIGRLGTRKRNFTLL 991
Db 981 LE-EDINKL-----KKTLDSEFLYNNKYKLEKLEKDKKRYKMOIKKLT---LL 1029
Qy 992 KLPLPIQSDLTJTTGSSSRP---LSAFRSGSGIFDIYVLPKSKNLEKLTAPLLKLV 1048
Db 1030 K-----EQLSEKLSLNNPKHVLONFVFFNK-----KKRAE----- 1062
Qy 1049 GVLACAFILVPSADAVDALKTACCLKGR---IELAKCIANPACANAVACLOTNNRDE 1106
Db 1062 -----IAETENLTENKILKHKYKGLVAYNNESSPLTSLSESIQTEDNVASL 1110
Qy 1107 TECQIKGDLFENSVDDENECAVSRKCVPRKSDLGFFRPDPSVLVONFNISDFNGKW 1166
Db 1111 E--NFKVLSKLEGLKLDNIN---LEKKRLSYLSGLHHLIA-ELKEVIKKNKY----- 1158
Qy 1167 YITSGINPTEDAFDQGLHEFHTEGDNKLVGNISMRIKTLDSGFFTRSAVQKVFOD----- 1222
Db 1158 ---TGNSPS-----ENNTDVNNALF---SYK-KFLPBGTDVAIVYSSGSDTLEQS 1201
Qy 1222 -PNOGVLNHNDEYLVHODDWYLLSKIEKNKPEDYIFVYVGRNDA-WDGYGAVVYTR 1279
Db 1202 QPKKPA-----STHVGAESNTITTSQNVDDVDVILVPIFGESEEDYDLGQVW--TG 1253
Qy 1280 SSYLPNSITPELEKAKKSGRDFSTFRDNTCGPREPLVRIETVEEGRIIVKEVEE 1339
Db 1254 EATPTPSVI-----DN-----ILSKTE--NEYVLVLYKPLAG 1282
Qy 1340 IEVEVEKEVEKVRTEMTLFORLAEFNEFKODEENFYRELSEKMEFLD 1389
Db 1283 VYNSLKQLE---NNMTFNWVKILNRFKRNENFNVLSEDLIPKXD 1329

RESULT 12
MSPL_PLAFW STANDARD: PRT: 1639 AA.
ID MSPL_PLAFW
AC 13-AUG-1987 (REL. 05, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P195).
GN MSP-1.
OS PLASMIDIUM FALCIPARUM (ISOLATE WELLCOME).
OC EURARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 8601435.
RA HOLDER A.A., LOCKYER M.J., ODINK K.G., SANDHU J.S., RIVEROS-MORENO V.,
RA NICHOLS S.C., HILLMAN Y., DAVEY L.S., TIZARD M.L.V., SCHWARZ R.T.,
RA FREEMAN R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites."
RT NATURE 317:270-273(1985).
RN [2]
RP REVISIONS.
RA HOLDER A.A.;
RA SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KD, 42
CC KD AND 19 KD ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

Cc -----
Cc EMBL: X02919; G5865;
Cc DR PTR: A24594; A24594.
Cc DR PFM: PFM00008; EGF; 1.
Cc KW MALAKIN; MEROZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
Cc TRANSMEMBRANE; GPI-ANCHOR.
Cc FT SIGNAL 1 19 POTENTIAL.
Cc FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
Cc FT CARBOHYD 116 116 POTENTIAL.
Cc FT CARBOHYD 268 268 POTENTIAL.
Cc FT CARBOHYD 764 764 POTENTIAL.
Cc FT CARBOHYD 768 768 POTENTIAL.
Cc FT CARBOHYD 783 783 POTENTIAL.
Cc FT CARBOHYD 844 844 POTENTIAL.
Cc FT CARBOHYD 920 920 POTENTIAL.
Cc FT CARBOHYD 964 964 POTENTIAL.
Cc FT CARBOHYD 1058 1058 POTENTIAL.
Cc FT CARBOHYD 1165 1165 POTENTIAL.
Cc FT CARBOHYD 1174 1174 POTENTIAL.
Cc FT CARBOHYD 1445 1445 POTENTIAL.
Cc FT CARBOHYD 1526 1526 POTENTIAL.
Cc FT SEQUENCE 1639 AA; 187618 MW; F0860D6A CRC32;

Query Match 1.8%; Score 134; DB 1; Length 1639;
Best Local Similarity 18.0%; Pred. No. 4.7;
Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

Qy 184 VOQNECAVSRKCVPRKSDGFFRPVDRNAVQNFMMKDFSGKWTITSLNPTFAFD 243
Db 249 ININELLSESKRTIDK-----NKNATKEEKKLQAOYDLS-----LYNK 290
Qy 244 QLEFFHENDKLVGNLTFRIRIKTLDSGFFTRSAVQTF--VQDP-----DLPGALYNHDE 295
Db 291 QLEEAH---NLISVLEKRIDILKKNENIKELDKINEKKNPPANGNPTLLDKKK 346
Qy 296 FLHYDDWYLLSQLEKNKPDY-----IYVYGRNDAMDYGSGVYTRSPLEPSII 349
Db 347 IEHEKEIKELIAKTIKFINDSLFTDPLEYLEYREKKNID--ISAKVETESTEPNE-Y 403
Qy 350 PNIQKAKSVGDFNENITTONSGPREPLVERLEKTAEEGKLLI-----KEAVEIEE 404
Db 404 PNCVTPPLY-NDINNALNELNSFG--DLINPDYIKPEPKNIYTDNEKKFLNEIKK 459
Qy 405 VEKEVEKVRDEMTLFORLAEFNEFKODEENFYRELSEKKEILNEL-----QMEAT 457
Db 460 IKLEKKKI-----ESDKSYEDRSKSLNDIKREYKELNELYDSKFNNDIDL 507
Qy 458 EYEKLFGR--ALPIRKLMAVAHHCFTSPCHDRIREFSSDDGIGRLGTRKRNFTLL 514
Db 508 NEKKMGKRYGYKVEKL-----THNTFASYENSKHN--EKLITRALKY--MEDYSLRNI 558
Qy 515 YVAKICSSRIIRYFPRKSPRICGLDSRGLDLSHGKLNLSPAHSINONVPRKSGC-- 572
Db 559 VVEKEL-----KTYK-----NLISITENEI--ETIVENIKKDEQLFEK 595
Qy 572 -----KFPDYALWMEKMGOFAPATAIVAIFILSVASKADAVALKTCGLKECRLE 625
Db 596 KITKDKNKPKDEKLEV---SDIVKVOQVLLM-----NKIDELKKTQLILNKVEIKH 645
Qy 625 -----LAKCISN-----PACAA-----NVACLOTNNRDET 651
Db 646 NIHPNSYKQENKQEPYLYIVKLELDKLYMPVEELINEEKKNITGOSQNSSEPT 705
Qy 652 EEOI-----KCGDLFENSVDDENECAVSRKCVPRKSDGFFRPDPSVLYQ-- 700
Db 706 EGETGQAITTKGQOQGSALG---DSVQAQOQDQKQAP-----PVPVYPEAKKQVP 756
Qy 700 -----KFDMDKDFSGKWFITRGLNPTDADFQGLHEFHTEGDNKLVGNLSKRI 745
Db 757 TTPAPVNNKTENYSKIDYLE-----KLYEFL-----NTSYIC 788
Qy 746 RTDGGFTTSVAVQKRVQDPKYPGLIYNHD--NEYVLVQDDWYLLSKVEN-----SPED 798

```

```

Db 789 H-----KY-----LVSHSTNKEKILKQ-----YKIKKESEKISSODPLD 824
QY 799 YIF-----YKKGKNDAMDGSGSV-----LYTRSAVL-----PSIIPHOATAQ 839
Db 825 LBNINONNIPVMT-----SMWDSLNNSLSOLFMEIYKEKVCNLYLKDKNDKIKNLEAK 880
QY 840 KVRGDNFTIKIDNTGCGPEP-PLVERLEKKVEGEFT-----LIKEVEE----- 884
Db 881 KV-----STVKTLSSSSMQLSLTPDCKREVSANDTSTLNNSLKLFENILSLGK 936
QY 884 -----EEVEKYRDEKVELFSLFEGFKELQDEENFLRELSEEMDLGK 931
Db 937 NIYQELIGKSSSENFYEKILKDSDFENESFTNFVSKADINSINDESKR-----K 989
QY 932 MEATEVEKLEGRALPIRKIMAVATHCFTSPCHDRIRFESSDGIQRLGTRRINGTELL 991
Db 990 LE-EDINKL-----KTLQLSFDLYNKYKLELEFKKKTVKRYKQIKRLT---LL 1038
QY 992 KILPPIQADLRTTGRSSRP---LSAFRSQFSKGFIDVPLPSKNEKELTAPLLKLV 1048
Db 1039 K-----DQLESKLSLNPNKPHVLQNFQVFENK-----KKEAE----- 1071
QY 1049 GYLACAFLLVPBADAVALKTCACILKGR---TELAKCIANPACANVAQCTCNRPDE 1106
Db 1071 -----IAETENTLENTKILKHYKGLVYKYNGBSSPLKTLSESIQTEDNYASL 1119
QY 1107 TECQIKGDLFENSVDENECVAVSRKCVPRKSDLGEPADPAPSVLQONNISQFNKW 1166
Db 1120 E--NKRVLKLEGLKLDNN---LEKKIISTYSSGLHHIIELEKVIKKNY----- 1167
QY 1167 YITSGINLPFDACQLHEFHTEGDNKLVGINSWRIKTLDSGFFPRAVQKRVQD----- 1222
Db 1167 ---TGNSSP-----ENNIDVNNALF---SYK-KFLPECTDVAIVVSSGSDTLEQS 1210
QY 1222 -PNQGVLYNHNDNEYLHVQDDWYIILSKIKENKPEYIIVYRGRDA-WDQYGAQVYTR 1279
Db 1211 QPKRA-----STHYGASNTITTSQNVDDVDVITVIFGSEEDYDLDGOV--TG 1262
QY 1280 SSVLNSIIPLELEKAASIGRDFSTFRTDNTCGEPALVRIEKTVEGERTIVKEVEE 1339
Db 1263 EAVTPSVI-----DN-----ILSKIE---NEVEVLKPLAG 1291
QY 1340 IEEVEKEVEKGRTEMFLFQRLAGFVEMLKODENFRELSKEMEFLD 1389
Db 1292 VYRSLKQDE---NNVMTFNVAVKDILNSRENKFNVLSDIPIYKD 1338

RESULT 13
MYSB_MESAU STANDARD: PRT: 1934 AA.
AC P13540: 060540;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH7
OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).
OC EDUAROTA; METAFOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; CRICETINAE; MESOCRICETUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FLB: TISSUE-LIVER;
RX MEDLINE: 95115033.
RA WANG R., SOLE M.-J., COKERMAN E., LIEM C.-C.;
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
heavy chain gene from Syrian hamster."
RL J. MOL. CELL. CARDIOL. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OF 962-1935 FROM N.A.
RX MEDLINE: 88247788.
RA JANDRESKI M.A., SOLE M.-J., LIEM C.-C.;
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin

```

```

RT heavy chain."
RL NUCLEIC ACIDS RES. 16:4737-4737(1988).
CC -I- FUNCTION: MUSCLE CONTRACTION.
CC -I- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -I- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -I- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEATITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -I- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEMOOSIN (LM)
CC AND 1 HEAVY MEMOOSIN (HM). IT CAN LATER BE SPLIT FURTHER INTO
CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -I- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L12104; G402372;
CC DR EMBL: X07273; G49641;
CC DR PIR: A28298; A28298.
CC HSSP: P08789; IAMD.
CC -----
KW MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;
KW ATP-BINDING; METHYLATION; ALKYLATION; MULTIGENE FAMILY.
KW DOMAIN 1 838
KW DOMAIN 839 1934
KW DOMAIN 779 801
KW NP_BIND 177 184
KW DOMAIN 654 676
KW DOMAIN 756 770
KW MOD_RES 128 128
KW MOD_RES 694 694
KW MOD_RES 704 704
KW CONFLICT 966 966
KW CONFLICT 978 978
KW CONFLICT 986 986
KW CONFLICT 1008 1014
KW CONFLICT 1057 1057
KW CONFLICT 1060 1060
KW CONFLICT 1095 1095
KW CONFLICT 1217 1217
KW CONFLICT 1271 1271
KW CONFLICT 1327 1327
KW CONFLICT 1358 1358
KW CONFLICT 1504 1504
KW CONFLICT 1537 1537
KW CONFLICT 1556 1556
KW SEQUENCE 1934 AA; 222928 MW; E586AA2F CRC32;

Query Match 1.8%; Score 133.5; DB 1; Length 1934;
Best Local Similarity 17.0%; Pred. No. 6.3;
Matches 251; Conservative 211; Mismatches 525; Indels 493; Gaps 63;

QY 188 NECAVSRKKCVPRKSDVGEFVPDRN-----AVQNFNMKDFSGKWITSGLPNT- 238
Db 390 NSADLKGCCHBRVAVGNEVYTKGQNVQVYALGALAKSYEKMFN--WAVTR-INTL 446
QY 238 -----PDADDCQLHF---FHMDKLVGNLWIRIKYTLDGGEFTTSAYQ 277
Db 447 ETKQPROYFLGLDIAGLEIDFNSFEOLCINFNEKLODFNNHMFVLDOEEYKKKGIE 506
QY 278 -TFVQ-----DPLDGLX-NHNEFLHYQDDW 303
Db 507 WTFIDFGMDLQACIDLIEKPRMISILEECMPKADMTFKAKLYDNHGLKSNKPKPR 566
QY 304 YILSQIENKPDQYIFVYRGRND---AMDGYGSGVYTRSPITLPSIIPNLQKAA-KS 358

```

Db 567 NKKGQOEK-----ESLVHACTVDYNTLIGW-----LQNKKDPLNTVYGLYQKSLK 615
 Qy 359 VGRDNNTITDNGSGPEPLVERLEKTAEGEKLIRKAVEI--EEVEKEVEKYADIE 416
 Db 616 LSNLFANI-----AGANAP--VDKGGKAKKSSSF-----QIVSVLHRENLKMTNLRST- 665
 Qy 417 MTLFQRLLEGEKELOODENFVRELKSEKE------LNLQLEAI--EYKELFGRA 466
 Db 665 -----HPHFVRCIIPNETKSPGMNDPLVMHQLRCNGVLEGIRICRKG 707
 Qy 467 LP-----IRKLMLAPHSNFIANHETIKYVGSGLPGHKREFSMGWDYGSIV 515
 Db 708 FPNRILYGFRORYRILNPAIPEGQFIDSKRGAELSSLDIDHNOYKFGHTKVFPAK 767
 Qy 516 VAKICSSRRIPYFRKSPRICGLDSRGLQJFSGKHNLSFAHS---INONVRKNSGCK 572
 Db 768 LIGLLEBMRDELSIITRI--QASRGLLSRMEFKLLEBRDLSLYIOWNI--RAFMGVK 824
 Qy 573 -FP-----KVALMYEKWQGFAPKTAIYAIILSVASKADAVDAKLT 613
 Db 825 NMPWMLKYEKIKPLKSAETEKEKATMK--EEFG-----VKDALEK 864
 Qy 614 CTCLKECHLELAKCISNPACANACIQTQNNRPDETECOIKGDLFENS- VDE 668
 Db 865 SEARRKELEKMYSLLOE---KNDLOLOVOAEDONLADAEERDQILTKNIOLEAKYKE 920
 Qy 669 FNE-----CAVSRK---KCVPRKSDVGFPPDPDSVLYVQFDMKDFSGWFI 713
 Db 921 MTERLEDEEMNAELTAKKRKLEDECESELRIDDELLETLAKVEKD----- 968
 Qy 714 RGLNPTFADFQOLHEFHTEENKLVNLSWRITPDDGFFTRSAVQKVFQDKIPGLIYN 773
 Db 968 -----HATEMK--KNLLEEMAGLD-----ELIAKLTKKEKALQEAHQ 1003
 Qy 774 HDNEYLYYODDWY--ILSKVE-----NSPEYIFVYKGRND---ANDGGSVLYTRS 823
 Db 1004 QALDDLOAEDEKVNLTJTKSKVLEQOVDELSLEOEKVRNDELRARAKLEGGDKLTQE 1063
 Qy 824 AVLPESIIPELOTAQKVQ-----RDE-----NFIKTNTCGPE-----PFLVERL 865
 Db 1064 STM-----DLENDKOODEKJLKKDFELNALNARIEQALGSLQKRLKELQARIEEL 1117
 Qy 866 EKKVEGERT-----IKVEEIEEVEEK----- 890
 Db 1118 EEBL--BAERKARAKVEKLKSDLSRELEISERLSEEGAGTSVOIEMNKKREAFOKMRD 1176
 Qy 890 -----VDKEVTLFSKLFEGFELQDEENFLBELSKEMENDVLDGLKMEAT 935
 Db 1177 LEBEATLOHEATTAALRRKKNADSVAELEQIDNLOKRVOKLEKESEFKLELDD---VTS 1232
 Qy 936 EVELKGRALPIRKLMVAVTHCTPSCHDRIRFESSDDIGRLGTRKRINGTFLKILP 995
 Db 1233 NMEDITAKANLEKM-----CET--LEDOUMNHR-----KAELQORVNDL 1273
 Qy 996 PIOSADLRTTGGRSSRPL---SAFRSGFSKGFIDYVLPESKNEKLELTAAPILLKLVGLA 1052
 Db 1273 TSOQAKIQTENGELSRLDEKELJISQLRSG---KLTYOOLEDLKROL----- 1319
 Qy 1053 CAPLIVSADAVDAKLTQACLLKGCRIELAKCIANACANAAACIQTQNNRPDETE--CQ 1110
 Db 1319 -----EEBVAKNTLALHLOSARDH-----CDLLRQYEELETKAE 1355
 Qy 1111 IKC-----GDLEFNSVADENECARVPRKSDLGEPAPDPDSVLYVONEN 1158
 Db 1356 LQCVLSTANSNEVQWKTETDAIQRIEELKEKKKLAQRLDAEF----- 1402
 Qy 1159 ISDENGWYIYSGLNPTFADFQOLHEFHTEGDNKLVGNISWRIKTLDSGFTRSVQKE 1218
 Db 1402 -----AVEAVNARCSSLEKTKHRLQNEIEDLWMD-----VERSNAAA 1439
 Qy 1219 VQDPNQGCVLYHNDNEYLHODDWYILSSIEKNEPEYIYVYTRGKNDAMDGGCAVYT 1278

Db 1440 ALDKQO-----RVEDKILAEWKQYEESSOSELSSOKE-----ARISST 1478
 Qy 1279 RSSVLPSSIIITLFE-----KAKSIGRDESFIFIRIDNTCGEPALVERIEKTEVEGERI 1332
 Db 1479 ELFLKNAVESELEHTEFRKNNLQEEISDITDLOSGTSGSIHELEIKRQLAENKE 1538
 Qy 1333 IKVEEIEEVEKEVEKAVRTEMTLFO-----PLAGFNELOKODENFVRELS----- 1382
 Db 1539 LOSALEEENSLHEBENIILRAQLEFNOIKAELTERLAKDEEMQAKNNHLRYVDSIQT 1598
 Qy 1382 -----KEEME-FLDEIKMEASEVEKLFKA 1405
 Db 1599 SUDAETRGRNALRVKMKMGDLNEMEIQLSHANRAA 1638
 RESULT 14
 LEF_BACAN
 ID LEF_BACAN STANDARD; PRT; 809 AA.
 AC P15917;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).
 GN LEF.
 OS BACILLUS ANTHRACIS.
 OC PLASMID PXOI.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 NC BACILLUS.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.
 RX MEDLINE: 90034185.
 RA BRAG T.S., ROBERTSON D.L.:
 RT "Nucleotide sequence and analysis of the lethal factor gene (lef)
 from *Bacillus anthracis*."
 RL GENE 81:45-54(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA LOWE J.:
 RL SUBMITTED (APR-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP ZINC-BINDING.
 RX MEDLINE: 95154659.
 RA KOCHT S.K., SCHIAVO G., MOCK M., MONTECUCO C.:
 RT "Zinc content of the *Bacillus anthracis* lethal factor."
 RL FEMS MICROBIOL. LETT. 124:343-348(1994).
 CC -1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
 CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
 CC DEATH. LF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED
 CC WITH PA, CAUSES DEATH. LF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO
 CC BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY
 CC FACILITATING THE INTERNALIZATION OF LF OR EF.
 CC -1- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT
 CC PROTEINS, A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN
 CC EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B. ANTHRACIS EF
 CC AND LF.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC
 CC METALLOPROTEASE).
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC CC
 CC EMBL: M29081; GI43144; -
 CC EMBL: M30210; GI43142; -
 CC PIR: J00032; J00032.
 CC PROSITE: PS00142; ZINC_PROTEASE: 1.
 CC HYDROLASE; METALLOPROTEASE; ZINC; TOXIN; SIGNAL; REPEAT; PLASMID.

FT SIGNAL 1 33
 FT CHAIN 34 809 LETHAL FACTOR.
 FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
 FT METAL 300 420 REPEATS.
 FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
 FT ACT SITE 720 720 POTENTIAL.
 FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
 SQ SEQUENCE 809 AA: 93786 MW: D81B6EBB CRC32;

Query Match 1.8%; Score 132; DB 1; Length 809;
 Best Local Similarity 17.2%; Pred. No. 2.4;
 Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 EENKLVGNLSWIRTPDGGFFTRSAVQKRVDPKYGILYND---NEELLYQDDMY--I 787
 DB 105 EMKKAIG---KIYVDGDTIKHISLEALSEKKRKIKDGLGKDALHEHYVAKGEYEPV 161
 QY 788 LSSKVENSPEDYIFVYKGRNDAMDGSGVLYTRSAPVLESIIPELQTAQKVGREDNT 847
 DB 162 L--VYQSSDYVENTERKALN-----VYELGKILSRILSKINOPYKFLDVINT 209
 QY 848 FIKTDNIGC-----PEPPLVERLEKKEVEGERTIK-- 879
 DB 210 IKNASDSDGDLFTNQLEKHEPTDFSEVFLLEONSNEVQEFKAPAYITDPQHRDVLQLY 269
 QY 879 -----EVEIEEVEKVRDKEVTLFSKLEFEGKE----- 908
 DB 270 APEAFNYMKFNEQELINLSLELKQDRMLSRLEKWKETIKQHOHWSDSLSEEGRLKKL 329
 QY 908 ---LORDEENFLRELSKEEMVDLGLKMEATEV---EKLGRALPI-----RKL 950
 DB 330 QAIPEPKDDIHSLSQEKELKRLQIDSSDFLSTEKEFLKQIDIRSLSEKEL 389
 QY 951 MAVVAHCFSPCHDRIRFSSDDGIGRLGITKRIKRIKNGFLKILPIIOSAD---LRTTG 1006
 DB 390 L-----NRIOVDSNP-----LSER--EKEFLKRLKLDIOPYINORLQDTG 429
 QY 1007 GRSSRPJ-----SAFRSGFSKGIETVLPSPKNELEKELAPPLKLVG 1049
 DB 430 GLIDSPSINLDRKQYKRDIONIDALLHQSIGSTLYNKIYLYEMKNINNLATL----- 484
 QY 1050 VLACAFILVPSADVAIDAKTACCLKGRICLANCIANPACANVACLQTCNNRPDETEC 1109
 DB 484 -----GADLYDST----- 492
 QY 1110 QIKGDLFENSVDI--FNECAVSRKKCVPRK---SDLGEPADPSVLYONENISDPNG 1164
 DB 492 -----DNTRKINGIFNEFKKPKYISSNYMIVDINERPALDNERLKWRIQLSPDTR 543
 QY 1165 KWYITSGINPTFPAFDCQLEHFTGDKLV--GNISWRITLDSGFTTSRQVQFVODP 1222
 DB 544 AGYLENG-----KLILQNRNIGLEIDVQ----- 567
 QY 1223 NQPGVLYHNDXYLHYQDDWYILSKTIENKPEYIFVYKGRNDAMQYGGAVVYTR--S 1280
 DB 567 -----IIKQSEKVIIRI--DAKVPSKIDTKIQE---AQININQEMKAKGLPXYTKLIT 617
 QY 1281 SVLPNSIIPLEKAKSISGRDFSTFIRTDNCGEPALVERIEKTEVEGE--RIIVKPV-- 1338
 DB 618 FNVHNRASNVESAYILLNEMKNIOSD-----LIKVIYIVLDGGRVFTDITL 669
 QY 1338 -----EIEIEVEKEVEKVGRTMTLLFORLAGFNELEKODENFVRE----- 1380
 DB 670 PNIAEQYTHODEIYEQVHSGLYVPESRSILLHPSKGV--ELRNDSEGFITHEGHAVDVY 728
 QY 1380 -----LSKEEME-----FLDEIKMEASEV 1398
 DB 729 AGYLLDKNQSDLYVTSNKKFIDIFKEBGSNL 758

ID Y109_YEAST STANDARD; PRT; 1679 AA.
 AC P40457;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 195.1 KD PROTEIN IN DNAA3-UBI1 INTERGENIC REGION.
 GN Y11149C.
 OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCYCETES; SACHAROMYCETALES;
 OC SACHAROMYCETACEAE; SACHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
 RA CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,
 RA GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
 RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
 RA RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
 RA WALSH S.V., WHITEHEAD S.;
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: 247047; G576197;
 DR EMBL: 238059; G557774;
 DR PIR: S48385; S48385.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 1679 AA: 195141 MW: 5897CD94 CRC32;

Query Match 1.7%; Score 130; DB 1; Length 1679;
 Best Local Similarity 17.9%; Pred. No. 8.3;
 Matches 210; Conservative 188; Mismatches 396; Indels 376; Gaps 59;

QY 379 IVERLE-----KTAEEBEKLLKEAVEIEEVEKVRDTEMTLFORLLSGFKELQ 431
 DB 536 LADKLEENYEQKQDKLQKVENOTIKEADAIIELE--NINAMETRIINLLREDSYKLLA 594
 QY 432 QDEENFVR-----ELSKKEKILNELQMEATEVEKLFGRALPIKRLR----- 474
 DB 595 STEENKANTNSVTSMEAREKKIRLELAELSTYVE---NSAIIQNLKRLLIYKKSQC 650
 QY 474 ---MALAPHSNF--LAN-----HETIKYVGSKLPGHRRSWMGMEVFFSGIVAKICSS 522
 DB 651 KKKTLLEDEENKGLAKKEKRLLEAIDHL---KALEKQKSW----- 691
 QY 523 RRIPIRFR--KSPRIICGLDLSRGLOLFSHGKHLNLPASHINQ--VPKNS-----CKKF 573
 DB 691 --VPSYIHYEKERASTELQSRIKISL--EYELISKKETASFIPTEESULTRDFEQCKE 747
 QY 574 PKDVAMWME-----KWQFATAIVAI-----FILVASKADAVALAKT 613
 DB 748 KKELOMRKESSEISHNENKMDSSKEGY--KAKIKLENNLERLRSLOSKIQIEISIRS 806
 QY 614 CTCLKECRLELAKCISNPACANVACLQTCNNRPDETECQIKGDLFENSVDIENFNECA 673
 DB 807 C-----KDSQKMAQ-----NIIDTEMKMK-----SLTLEISNKE 837
 QY 674 VSRKKCVPRKSDVGDPEVPPDPSVLVQKFDKMGSKWFTIRG---LNPJ---FPAFD 724
 DB 838 TTIKRL---SSEIENL---DKELRKTRKFOYK-----FLDQNSDASLTLEPTIKRELEIQ 885
 QY 725 COLHEFHTENKKLGVNLSWIRTPDGGFFTRSAVQKRVDPKYGILYNDNRYLYQDD 784
 DB 886 VOLKDANS-----QIQAEEIISNENALI----- 911
 QY 785 WYLLSKVENSPEDYIFVYKGRNDAM-----DGYGGSVLYTRSAPVLESIIPELQ 835

RESULT 15
 Y109_YEAST

```
Db 911 --ELNNEIAKTKENYDAIELEKKEKWARBEDLSRLRGELGI-----BALQPKL 959
QY 836 TAAQKVGDPNFTKTDTCGPEPLVERLKEKVEGER--TIT-----KEVEIEEEVE 888
Db 960 BGALH-----FVOOSEKLNE--VERIQKMEIEKEMSTIVOLCKKEMSOYQSTMK 1009
QY 889 KVRD-----KEVT-LFSKLFEGFELORDEENFLRELSKREMDVLQK 931
Db 1010 ENKJDELIVLEKMAADCAELITKTSLSADLDLKHKKMEKADYERELISNIE 1069
QY 932 M-EATEVEKLGKALPIKMLAVATH-----CFTSPCHDRIRFFSSDDIGRLG 979
Db 1070 QTESLAVE---NSVLEKVDYDTAANNQDKHLKLVSLFSLRHER----- 1112
QY 980 ITRKRINTFLKLILPIQSADLR--TTGGRSSRPLSAFRSGSGIFDIYPLPSKNELKE 1038
Db 1112 -----NSLETYKLTCTCKRELAFVKOKNDSLEKTINDLORTQTLSE-KE 1152
QY 1039 LTAPELLKLVGLVACAPLIVPSADAVDALKTCAQLKGR:ELAKCIANPACANVACLQ 1098
Db 1153 -----YQCSAVIIDEFKDIKEYTQVAILKENNAILOKSLKN-VTEKNREIYK 1199
QY 1099 TCNNRPDETECOIKCGDLEFNSVDEFEKAV-----SRKKCVPRKSDLG--EFPADPS 1151
Db 1200 QLNDRQOEISRLQR--DLIQTKEQVSINSNKLIVYEESEMEQCKORYODLSQCKDAQKD 1257
QY 1152 VLVOFNINISDENGKWTITSGINPTFEDAFDOLHFEHTEGDNKLVGNISWRIKITDSGFTT 1211
Db 1258 IEKLTNEISDLKGR--LSSAENANADL-----ENKEN-----RLK----- 1291
QY 1212 RSAVQKFEVODPNQPGVLYNHUNEY-----LHYODDWYI-LSSKI--ENKPEDYI 1257
Db 1291 KOAHEKLDASKKQQAALTNELNEIKAIKDLQDLHFENAKVIDLDTKKAHELQSEDY- 1350
QY 1258 FVYTRGRDAMDYGAVVITRSSVLPNSIIPLELEKAKSIGRDSFTFIRDTNCGPEPA 1317
Db 1350 -----SRDEKDY-----RTLMEIE-----SLKELQIF-----KTANSSD 1383
QY 1318 LVERIEKTV-EGERIIVKEVEIEEVEKEVEKVGRTMTLFORLAGFNE-LKODEEN 1375
Db 1384 AFEKLVNMEKEKDRITIDERTKEFEKKLOETLNKSTSSAEYISKDITETLKKEWLKEYEDE 1443
QY 1376 FVRELSEMEFEIDELIKWEASE-VEKLEK 1404
Db 1444 TLRRKEAEENLKKRIRLPSEERIQTISK 1473
```

Search completed: October 14, 1999, 03:57:50
Job time: 2383 sec

•
•
•
•

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 1999, 22:44:45 : Search time 44.23 seconds
(without alignments)
1964.721 Million cell updates/sec

Title: US-09-075-375-4
Sequence: 1 MALSHTVFLCKEALNLYA.....MEASEVKLFGKALPIKRV 1412

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database:

SPREMBL.10:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2557	34.1	478	10	Q40593	Q40593 nicotiana t
2	2508	33.5	473	10	Q40251	Q40251 lactuca sat
3	2430	32.4	462	10	Q39249	Q39249 arabidopsis
4	194.5	2.6	2269	5	Q26223	Q26223 plasmodium
5	167.5	2.2	2401	5	Q26216	Q26216 plasmodium
6	156	2.1	1365	2	Q49525	Q49525 mycoplasma
7	156	2.1	1933	13	Q90337	Q90337 cyprinus ca
8	155.5	2.1	886	1	Q29230	Q29230 archaeglob
9	151	2.0	1939	5	Q23622	Q23622 plasmodium
10	147	2.0	3113	4	Q13246	Q13246 homo sapien
11	146	1.9	2748	3	Q03767	Q03767 saccharomyc
12	144.5	1.9	839	5	Q26024	Q26024 plasmodium
13	144	1.9	2166	5	Q51465	Q51465 borrelia bu
14	143.5	1.9	1558	5	Q96275	Q96275 plasmodium
15	142.5	1.9	1819	2	Q924V0	Q924V0 helicobacte
16	142.5	1.9	1199	5	P91349	P91349 caenorhabdi
17	142.5	1.9	5105	2	Q61201	Q61201 caenorhabdi
18	141	1.9	1002	5	Q65883	Q65883 aquifex aeo
19	140.5	1.9	1302	2	Q49547	Q49547 mycoplasma
20	140	1.9	800	1	Q59066	Q59066 metanococc
21	139.5	1.9	1025	3	Q12176	Q12176 saccharomyc
22	137	1.8	1676	6	Q00756	Q00756 oryctolagus
23	137	1.8	1676	10	Q23332	Q23332 arabidopsis
24	136.5	1.8	3724	5	Q77320	Q77320 plasmodium
25	136	1.8	1127	12	Q9YVT6	Q9YVT6 melanoplus
26	135.5	1.8	1312	4	Q92878	Q92878 homo sapien
27	134.5	1.8	946	2	P70888	P70888 bacteroides
28	134	1.8	978	2	Q67124	Q67124 aquifex aeo
29	134	1.8	620	4	Q43663	Q43663 homo sapien

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	478 AA.
Q40593	1			
ID	Q40593			
AC	Q40593			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DI	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-JAN-1999 (TREMBLrel. 09, Last annotation update)			
DE	VIOUXANTHIN DE-EPOXIDASE PRECURSOR.			
GN	TVEL.			
OS	Nicotiana tabacum (Common tobacco).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;			
OC	Nicotiana.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-XANTHI; TISUE-LEAF;			
RA	BUGOS R.C., YAMAMOTO H.Y.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL, U34817; AAC50031.1; .			
DR	MEDEL; 9222; Nicotiana glauca; L.			
KW	Transit peptide.			
FT	TRANSIT 1 134 POTENTIAL.			
FT	CHAIN 135 478 POTENTIAL.			
SQ	SEQUENCE 478 AA; 54561 MW; 9882AD42 CRC32;			

Query Match 34.1%; Score 2557; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 2.5e-139;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	474	MALAHNSFLANHEITIKYVSGSKLPGRKFSWGMEDYIGSTIVAKICSSRIIPYFKSP	533
DB	1	MALAHNSFLANHEITIKYVSGSKLPGRKFSWGMEDYIGSTIVAKICSSRIIPYFKSP	60
QY	534	RICCGDLSRGLTFSHGKHLNPAHSINQNPVKGSGCKPKDVALMWEKGFAXTAI	593
DB	61	RICCGDLSRGLTFSHGKHLNPAHSINQNPVKGSGCKPKDVALMWEKGFAXTAI	120
QY	594	VAIFLISVASKADAVDAKTCTCLKECRLEFLAKCISNPACAAVACIQTCCNNRPDETEC	653
DB	121	VAIFLISVASKADAVDAKTCTCLKECRLEFLAKCISNPACAAVACIQTCCNNRPDETEC	180
QY	654	QIKCGDLEPNSVDFEPCAVSRKCVPRKSDVGFPPDPDSVLYQKDKMDQFSGKFEIT	713
DB	181	QIKCGDLEPNSVDFEPCAVSRKCVPRKSDVGFPPDPDSVLYQKDKMDQFSGKFEIT	240
QY	714	RGUNLPEDAFDQLEHFEHTEENKLVGNLSWRIRPDGFFTRSAVQKVPQPKPGILYN	773
DB	241	RGUNLPEDAFDQLEHFEHTEENKLVGNLSWRIRPDGFFTRSAVQKVPQPKPGILYN	300
QY	774	HONEYLLYQDDVYILSKSVENSPEDYIFVYKGRNDAMDYGGSVLYTRSAVLPESITPE	833

```

Db 301 HDNELYQDDWYLLSSVENSPEDDYLFVYKGRNDAMDGGSVLTKRSNVLPESTIPE 360
QY 834 LQIAAQKVGRODNFTIKTNTGCEPPLVERLEKKEVEGERTIIKEVEEIEEVEKVRD 893
Db 361 LQIAAQKVGRODNFTIKTNTGCEPPLVERLEKKEVEGERTIIKEVEEIEEVEKVRD 420
QY 894 EVTLFSKLFEGFKELORNEENFLRELSEKEMDVLDGLKMTAEVKEKLFGRALPIRL 950
Db 421 EVTLFSKLFEGFKELORNEENFLRELSEKEMDVLDGLKMTAEVKEKLFGRALPIRL 477

RESULT 2
ID 040251 PRELIMINARY; PRT; 473 AA.
AC 040251;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN VDEL.
OS Lactuca sativa (Garden lettuce).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asteridae; euasterids II; Asterales; Asteraceae;
OC Lactuca.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ROMAINE;
RX MEDLINE; 96270536.
RA BUGOS R.C., YAMAMOTO H.Y.;
RT "Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
RT and expression in Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:6320-6325(1996).
DR EMBL: U31462; AAC49373.1; -.
DR MENDEL; 8691; LACsa; Vdel; 1.
KW Transist peptide.
FT TRANSIT 1 125 POTENTIAL.
FT CHAIN 126 473 VIOLAXANTHIN DE-EPOXIDASE.
SO SEQUENCE 473 AA; 54447 MW; 102E7001 CRC32;

```

Query Match 33.5%; Score 2508; DB 10; Length 473;
 Best Local Similarity 100.0%; Pred. No. 1.6e-136;
 Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MALSLHTVFLCKEALNLYARSPCNEFRHSQGPNTIIMKIRSNNGYFNSRLFTSYK 60
Db 1 MALSLHTVFLCKEALNLYARSPCNEFRHSQGPNTIIMKIRSNNGYFNSRLFTSYK 60
QY 61 TTSFSDSHCKDKSQICSDTSEFEEIORFLDKRGMTLLEKORFOIOLAIYVCFFVY 120
Db 61 TTSFSDSHCKDKSQICSDTSEFEEIORFLDKRGMTLLEKORFOIOLAIYVCFFVY 120
QY 121 PRVDAYDALTKCQLCKECRIELAKCIANSCAANVACIOTCNRPDETECOIKGDEFE 180
Db 121 PRVDAYDALTKCQLCKECRIELAKCIANSCAANVACIOTCNRPDETECOIKGDEFE 180
QY 121 PRVDAYDALTKCQLCKECRIELAKCIANSCAANVACIOTCNRPDETECOIKGDEFE 180
Db 121 PRVDAYDALTKCQLCKECRIELAKCIANSCAANVACIOTCNRPDETECOIKGDEFE 180
QY 181 NSVVDQNECAVSRKCKVPRKSDVGEPPVDRAVONNENMKDFSSKWTITSLANTFPA 240
Db 181 NSVVDQNECAVSRKCKVPRKSDVGEPPVDRAVONNENMKDFSSKWTITSLANTFPA 240
QY 181 NSVVDQNECAVSRKCKVPRKSDVGEPPVDRAVONNENMKDFSSKWTITSLANTFPA 240
Db 181 NSVVDQNECAVSRKCKVPRKSDVGEPPVDRAVONNENMKDFSSKWTITSLANTFPA 240
QY 241 FDCQLHEFHENDKLVNLTWRITKTLDDGFFTSASVOTFVQDDPLDGLALYNHNDNEFLHQ 300
Db 241 FDCQLHEFHENDKLVNLTWRITKTLDDGFFTSASVOTFVQDDPLDGLALYNHNDNEFLHQ 300
QY 301 DDMYIILSSQIENKPDYIFVYVYGRNDAMDGGGSVYVTRSPLEPSIIPNLQAKASVG 360
Db 301 DDMYIILSSQIENKPDYIFVYVYGRNDAMDGGGSVYVTRSPLEPSIIPNLQAKASVG 360
QY 361 RDNFNFTTNSCGPEPLVERLEKTAEGEKLIIKEAVEIEEVEKVRDENTLIF 420
Db 361 RDNFNFTTNSCGPEPLVERLEKTAEGEKLIIKEAVEIEEVEKVRDENTLIF 420

```

```

QY 421 ORLLEGFKELQDEENFVRELSEKKEKELINELQDEATEVEKLFGRALPIRL 473
Db 421 ORLLEGFKELQDEENFVRELSEKKEKELINELQDEATEVEKLFGRALPIRL 473

RESULT 3
ID 039249 PRELIMINARY; PRT; 462 AA.
AC 039249;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN VDEL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA BUGOS R.C., YAMAMOTO H.Y.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEMAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURZ D., OJI O., SHEN Y.K., TORUMI M.,
RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THEOLOGIS A.,
RA ECKER J.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: U41433; AAC50032.1; -.
DR EMBL: AC003981; AAC14029.1; -.
DR MENDEL; 6341; ARAtb; Vdel; 1.
KW Signal.
FT SIGNAL 1 113 POTENTIAL.
FT CHAIN 114 462 VIOLAXANTHIN DE-EPOXIDASE.
SO SEQUENCE 462 AA; 52017 MW; 37FB33BA CRC32;

```

Query Match 32.4%; Score 2430; DB 10; Length 462;
 Best Local Similarity 100.0%; Pred. No. 4.5e-132;
 Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 951 MAVATHCFTSPCHDIRIFFSSDDIGRLGTRKRINGTFLKTLPPIOSADLRTTGGRSS 1010
Db 1 MAVATHCFTSPCHDIRIFFSSDDIGRLGTRKRINGTFLKTLPPIOSADLRTTGGRSS 1010
QY 1011 RPLSAFSGSGKGFIDVPLPSKNEKELTAPLLKLYGLACAFILVPSADAVDAKTC 1070
Db 1011 RPLSAFSGSGKGFIDVPLPSKNEKELTAPLLKLYGLACAFILVPSADAVDAKTC 1070
QY 1071 ACLKGRIELAKCIANPACAAVACIOTCNRPDETECOIKGDDFENSVDVEFNECAV 1130
Db 1071 ACLKGRIELAKCIANPACAAVACIOTCNRPDETECOIKGDDFENSVDVEFNECAV 1130
QY 121 ACLKGRIELAKCIANPACAAVACIOTCNRPDETECOIKGDDFENSVDVEFNECAV 180
Db 121 ACLKGRIELAKCIANPACAAVACIOTCNRPDETECOIKGDDFENSVDVEFNECAV 180
QY 1131 SRKCCVPRKSDLGEPAPDPSVLVONENISDENGKWTITSGLNPTEDAFDCQLHEPTEG 1190
Db 1131 SRKCCVPRKSDLGEPAPDPSVLVONENISDENGKWTITSGLNPTEDAFDCQLHEPTEG 1190
QY 181 SRKCCVPRKSDLGEPAPDPSVLVONENISDENGKWTITSGLNPTEDAFDCQLHEPTEG 240
Db 181 SRKCCVPRKSDLGEPAPDPSVLVONENISDENGKWTITSGLNPTEDAFDCQLHEPTEG 240
QY 1191 DNKLAVGISMRKIKTLDSGFTSRSAVOKFVODPNOPGVLYHNDNEYLHYDDWYIILSSKIE 1250
Db 1191 DNKLAVGISMRKIKTLDSGFTSRSAVOKFVODPNOPGVLYHNDNEYLHYDDWYIILSSKIE 1250
QY 241 DNKLAVGISMRKIKTLDSGFTSRSAVOKFVODPNOPGVLYHNDNEYLHYDDWYIILSSKIE 300
Db 241 DNKLAVGISMRKIKTLDSGFTSRSAVOKFVODPNOPGVLYHNDNEYLHYDDWYIILSSKIE 300
QY 1251 NKPEDYIFVYVYGRNDAMDGGGSVYVTRSSVLPNSIIPLEKAAASIGRDEFTFRTDN 1310
Db 1251 NKPEDYIFVYVYGRNDAMDGGGSVYVTRSSVLPNSIIPLEKAAASIGRDEFTFRTDN 1310
QY 301 NKPEDYIFVYVYGRNDAMDGGGSVYVTRSSVLPNSIIPLEKAAASIGRDEFTFRTDN 360
Db 301 NKPEDYIFVYVYGRNDAMDGGGSVYVTRSSVLPNSIIPLEKAAASIGRDEFTFRTDN 360
QY 1311 TCGPEPALVERIKKYVEGERIIIVKEVEEIEEVEKVRGTEMTLRFQRLAEGFNEIK 1370
Db 1311 TCGPEPALVERIKKYVEGERIIIVKEVEEIEEVEKVRGTEMTLRFQRLAEGFNEIK 1370
QY 361 TCGPEPALVERIKKYVEGERIIIVKEVEEIEEVEKVRGTEMTLRFQRLAEGFNEIK 420
Db 361 TCGPEPALVERIKKYVEGERIIIVKEVEEIEEVEKVRGTEMTLRFQRLAEGFNEIK 420

```


QY 1371 ODEENFVRELSEKEMFLEDEIKEMASEVEKLFGLALPIRKVR 1412
 DB 421 ODEENFVRELSEKEMFLEDEIKEMASEVEKLFGLALPIRKVR 462

RESULT 4
 026223 PRELIMINARY: PRT: 2269 AA.

AC 026223:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE RHOPTRY PROTEIN.
 OS Plasmodium berghei yoelli.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YM;
 RX MEDLINE: 95021522.
 RA KEEN J., SINHA K., BROWN K., HOLDER A.;
 RT "A gene coding for a high-molecular mass rhoptry protein of
 PL Plasmodium yoelli."
 RL Mol. Biochem. Parasitol. 65:171-177(1994).
 DR EMBL: L27838; AAA21304.1;
 SQ SEQUENCE 2269 AA; 265158 MW; E0A79FA6 CRC32;

Query Match 2.6%; Score 194.5; DB 5; Length 2269;
 Best Local Similarity 18.4%; Pred. No. 0.0069;
 Matches 255; Conservative 213; Mismatches 436; Indels 483; Gaps 65;

QY 304 YIIISQIENKPDYIFVYRGRNDAMDGYGGSYTRSR-----PLUPESIIPNLO--KA 355
 DB 34 YITISQIENKLN--VSTIPBGR-----EGFTSSDLAKSEKTKLETITELTSNETIVL 87
 QY 356 AKSVGRDFNNFITTNSCGEPPLVERLEKTAEGEKL-----LIKEAVEI 401
 DB 88 EKREIRELFKKY-----LDEBAERKYLEGLKLEKIKIDIAKIEVKNTVEL 135
 QY 402 EEEVER-----EVEKVRDTEMTLFQRLLEGKELQO--DEENFVRELSEK 446
 DB 136 KKEIEKNNAYIDELANOSPYPKVTGYIENKNTIYNTIKSYFDQIYEDIDTFEYNELSIYK 195
 QY 447 E-IINELQMEATEVEKLFGRALPI--RKLMLALPSNFEIANHET-----IKYYV 493
 DB 196 EDPIIDIE-DKIKLENIKSKIDNVYDKIQKMEIETVKSHTNINLETNNKLPNTILETKKI 254
 QY 494 GSKLPGRKFSWGMEDYFGSIYVAKICSSRRIPRYFRKSPRICGLDSRGLQLEFSH--- 550
 DB 255 YDEI--SKELINKMLEDFKNK---EKELSKNISDYDKRPEOL--SEKSKMLFIRNHVNSQ 307
 QY 550 -----GKHNLSPAHSHINQNP-----KNSGCKEPKQV 577
 DB 308 TNVDNTKEEBAKQYDKNSNEHMTIPTNEDEISKIISEVTKMDELLSKVNTYIDNKKY 367
 QY 578 ALAWMEKQGFAKTAIYAIIFILSVAKADAVDALKTCTCLKECRLELAKICSPACAA 637
 DB 368 KEIYNSEHSOFT-----LTDKIKAEVSDK-----ELKR----- 397
 QY 638 VACLQTCNNRP--DETQCOIKCGDLFEN---SVDENECAVSRKKCVPRKSD----- 666
 DB 397 --CQSFNDKSLINETKNSIE--KEYONINTLKKVDEYIKVCKSTKESITKFSKOTIL 452
 QY 686 -----VGDFPVDPSTLVOKFDMKDESGKWFITRGNTPTFDFAFDQLHEFHTEE- 735
 DB 453 KDMLNQIKTYKETSIDS--YIEKFE-QILTGK--QTKLENKTFEFSJNNEHANNEL 507
 QY 735 ---NKLWNLSMRLRTPDGGFTFSAVOKFVOPKYPGLIYNHNDNYLLYODWYLLSS 790
 DB 508 IKYSYDLKANGINDENNLYNOFTK--EKTFNDIKENKT--HINDEI-----S 552
 QY 791 KVENSPEDYIF-----VYKGRNDAMD-----GY 814

DB 553 KIEIKIHASIIYINISEETEREIGINIESLNTKYVEKVENYINLKIKEKLHYDSDEGK 612
 QY 815 GGSVLYTRS-----AY-----LPESIPELOTAOKYGRDFNFI 849
 DB 613 EGNKIYTDKIK:INDIMAVSQIDQHINGLDIDQKSSSEYSEMEQJINKLEKNSNTEI 672
 QY 850 KIDNTCG--PEPPLVERLEK--VEGERTIIEVEEIEE--VEKYRDEVILFSKL 901
 DB 673 SNQVSEIKKKKQIIVIKIDKDKKNITVEIKILSEISKIEKDNSTLEKYKDINLSYGQML 732
 QY 902 PEGKKEIQRDEENFLRLSKSEM---DVLDDG:KMATVEYKLFGLALPIRKIMAN--AT 955
 DB 733 GNLFLE-QIDEEKKAKENTIKSMAYIDDIDNTKRSQLEIETMDIKMDINKMEALKTS 791
 QY 956 HCFISPCHDRIRPSSDDGIGITRKRINGTFLKLIPPIOSADLRTGGRSSRPLA 1015
 DB 792 HDDDKRCHDKSK--NHKENISDIYDKSSKITIODE-----SRESIDINKNLOKNSVE 842
 QY 1016 FRSGFS-----KGFIDVPLPSKNELKEL-----TAP 1042
 DB 843 SQNHNSDINOCNEVANINYLKL--NKIKIIDKVEYTSIEERKKNKINDELNANSEK 899
 QY 1043 LLLKLVGLACAFIIVSADAVDALKTCACLLKGCRLELAKCIAN--PCANVACLO 1099
 DB 900 VIKRIEGLS-----LKECRSKINSTJDDKDIDECIKINIVLKK 938
 QY 1100 CNRPDETEC--QIKCGDLFENSVDENEC----- 1129
 DB 939 -NINLEETINITHKKNAEYKIVLSFNINIEADNKSQYILEIKKNNGINDHDYNIKEL 997
 QY 1129 -----AVSRKKCVPRKSDGEFPADPSVLYONFINSDFNGKWITSGLNPT 1175
 DB 998 KSHDKDSNGYGTAEQQKKAIOKKKELEFQYKEVEYVLKKNYAAVELKKN----- 1048
 QY 1176 FDADCOQHEHT:EGDNKLVANISMRIR-----TDSGFTFRAVQKFFQDPNPGVLY 1229
 DB 1048 --FD-----KTKNDKSK--QIKKEIKDAHNYCTLESK-----KSKKNME----- 1083
 QY 1230 NHDNELYHODQWYILSSKIEKPEDEYIFVYRGRNDAMDGYGAVYTRSSVLP----- 1285
 DB 1083 -IKNEKIHIED--VANNDKSNK-----ALISIVSVEPFTKI 1118
 QY 1285 --NSI-----IPELEKAASIGRDFSTIFRINDTCGPPALVERIKTEVEGER 1331
 DB 1119 IKINEIRTKSDCLKETNDKQJLSIDIOETKLIEN--GKOLKTLLELLESLKROK- 1176
 QY 1332 IIVKVEEIEEVEVEKVGRTMTLFOR-----LAEGFNLEKQDEENFY--RELS 1381
 DB 1176 --KNIEDQKRELDEVNSKIKNIENTVNOHKKNYEIGIVEIKINELIATKTKNOJESTKELI 1232
 QY 1382 KEEMFEL 1388
 DB 1233 KPTIOHI 1233

RESULT 5
 026216 PRELIMINARY: PRT: 2401 AA.

AC 026216:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE RHOPTRY PROTEIN (FRAGMENT).
 OS Plasmodium berghei yoelli.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YM;
 RX MEDLINE: 97077455.
 RA SINHA K.A., KEEN J.K., OGUN S.A., HOLDER A.A.;
 RT "Comparison of two members of a multigene family coding for
 high-molecular mass rhoptry proteins of Plasmodium yoelli."
 RL Mol. Biochem. Parasitol. 76:329-332(1996).

DR EMBL: 036927; AAB41263.1; -
 FT NON_TER 1
 SO SEQUENCE 2401 AA; 281980 MW; 39A40760 CRC32;

Query Match 2.2%; Score 167.5; DB 5; Length 2401;
 Best Local Similarity 16.7%; Pred. No. 0.26;
 Matches 260; Conservative 274; Mismatches 498; Indels 525; Gaps 71;

QY 23 PCRERHRSQPPPTNIMK--IRSNNGYFNSRFLFTSKTSSFDSSGCKKSOICSID 80
 DB 453 PPNEMKYQ--KSEIEIKIKDEFLSKVKNRYNDFD--KVVKEKVESEHNKFTELTKIKTE 508
 QY 81 TSFEIQRFDLKRGMFLILEKQWOFIQLAIVLCFTVIVPRVADAKTACGLKRCR 140
 DB 509 VSDEEIKKYEKNKNDKSLINERKKSIEEYONIN--LKKVD--DIYKVC----- 556
 QY 141 IELAKCIANPSCAANVACIOTCNR---PDTECOIKC-----GDLFNSYVD 185
 DB 556 LNTNELITN---CHNKOTLTKDKLNQNIKTIKETSINDKIYTDKFNILTD 603
 QY 186 Q-----FNECAVSRKKCV---PRKSDVGEFPPVRNNAVQNNKDKSGKMY 223
 DB 604 KTELETKFTGSLNHNHSENNKELLTFYDLKANLGR--NKENMLYKQFNEKEKAVE-- 659
 QY 230 ITSGNLTFDAFDCCQLEHFMENDKLVGNLTWRIKTLDGGEFTSRSAVQTFVDDPLGAL 289
 DB 659 -----DIKKKNVDINKIVSNIEITIT-----SI 682
 QY 290 YMHND-----ELHYQDDWYLS-----SOIENKPDYIFVYVYGRNDAMD 330
 DB 683 YINEDTENEICKSTELLTK--VLEKVKANVTNLEIKELKDYQFQDF----- 731
 QY 331 GYGGSYIYRSPTLBPESIIPLNLOKAASVGRDENNFITDSCGPEPLVRLKTAEG 390
 DB 731 GKEKNIKPYDEKIKINDITLQKIDKSEITETIKKXSENHIDIKQIDRLKVV--PN 788
 QY 391 EKLILKEAVEIEEVEKEVEKVRDTEMLFO--RLLEGFELOODEEN----- 437
 DB 789 KTFMNEDEPREIEKKIENIYEKI--DKKNIYKIDKLNEIKSEINDKTSLEKLNINISY 847
 QY 437 -----FVRELSEKEKEI-----LNLQMEATEVEKLPGRALPT-----RK 471
 DB 848 GKSLENTLQOIDEKKAHEHTIKAMEAYIDDLNKKKSQSEIKEMNINIMDKIDHKE 907
 QY 472 LEMALAPHSNLANHETIKYVGSKLPGHKRPS-----MGWEDYFGSI-----VYAK 518
 DB 908 KMALNISHDDYKIYHTSKNH--EEKISDIRKNSLKIIDFSEESYINDIKKELEKNVLES 966
 QY 519 ICSSSRIRYFRKSPRICCGDSRGLQFSGKHNLASHSINONVPRKNSGCKFPKQVA 578
 DB 967 ONNNTDIOYLSKIENT--NIILKLNKIKKIIDK----- 999
 QY 579 LMVEKMGQFAKIALVAIFILSVASKADAVDALKTCTL-----LKCRPLELAKCISN-- 632
 DB 999 --VKEYTDEIKN-----NKKINAEISSEKIIITOLKNSSLKCEQSKISTIDNY 1048
 QY 632 -PACANVACIOTC-----NNRPETECOIKGDLFENSVDDEFNECAVSRKKCVPRKSD 685
 DB 1049 VSEIKINTNTKIYIVNEKN-----INTFYKNA--EYVNO----- 1083
 QY 686 VGDFPVPSPVLQKFDKDKSGKWFITRGINPTFDFADQCLHEFHTEENKVLGNLSWRI 745
 DB 1083 -----VSLNFNNTIEADTKSOYL-----NITKNGNTNMDNYI 1117
 QY 746 RTPDGGEFTSRSAVQFVODPKYPGILYNHDEYLLYODDWYILSSKYENSPEDYIFVYK 805
 DB 1118 KE-----LKEHKKSNNVYKDE-----ACKNQF-----TK 1142
 QY 806 GRNDAMGIGGSVLYTSAYLPESITPELOTAQOKVGRDENTFKIDNTGPEPPLEVL 865
 DB 1143 KKNLEFKYEBOEV--TVLNNKYAVELK-----NKEFDKTN-----YSEQI 1181

QY 866 EKVVESEBRTIIEVEIEEVEKVRDKEVTLFSKLEFEGFKLQRDENFLRELKSEMD 925
 DB 1182 IKELIDAHNTFTSQADSEKKEKNEIKNEQIRI-----EDEVAKNNSKNA 1226
 QY 926 VLD-GLKMEATEVEKLEGRALPIRKLMVAATHCF--TSPCHDIRFFSSDDGIGRIGTIR 982
 DB 1227 ILDQLSVPEPKI-----KFLKIKDLRTKSDCLKETKDIETKISMLSIDQETKL-IGN 1280
 QY 983 KRINGFLKILPPIQASADLRTTGGRSSRPLSAFRSGFKGJFDIPLPSKNEKLELTAP 1042
 DB 1281 KNILNT-LEKLSLSKNO-----KKNIED-----QKHEDEVANSK 1314
 QY 1043 L-----LTKGVLAACAPLVPASDAVDALKTACILKGRICELACI 1085
 DB 1315 IKNIESNVNOKKKNYELGIEYKINEIAKA-----NKDQIESQOKLI-----IPTIKNL 1362
 QY 1086 ANPACANVACIOTCNRPD--ETECQ-----IKGDL-----FENSV 1121
 DB 1363 ISPFKANDLEGIDITNNNLKRYNTEANNIYEFKTSIDLITHTLEYVSKETIYEOIKNR 1422
 QY 1122 VDEFNECAVSRKCVPRKSDLGEPAPDSVLYQNF--NISDFNGKW--YITSGLNPTE 1176
 DB 1423 ITAQNELITNIKVNRAKSYLDDIEANEDRIYTHPKNKLNDVNDKFTNEY--SKVNGGF 1480
 QY 1177 DAFDCQLEHFEHTGDKKLIVGNISMRIKITLDSGFFTSRSAVQTFVODPNQGVLYNHNDNYL 1236
 DB 1481 DNISNSINNVKSTDENLNLTINOTKEM--YANIVSKRY-----SYK 1522
 QY 1237 HYODDWYILSKLENKPEDEYFVYVYGRNDAMDYGGAVVYRSSVLPSPISPELEKAAK 1296
 DB 1523 YEAMENIFINPKLAN-----SLNIQIKSSSGIDLFPNNI--ALIPYIDSOKE 1568
 QY 1297 SIGRDESTIRIDNTGCPPALVERIEKTEVEGERI-----YKVEEIEEVEKE 1347
 DB 1569 ---DILFETL-----PSP-----EKTSEYTYTISISYNTLIDLKRSQELQKREQQA 1611
 QY 1348 VEKVGRTENTLFORLAEGFNEIKODEENFVRELSEKMEFEIDELM--EASEVYL 1401
 DB 1612 LNLIL--FEKRLHDVKQVATNEIK-----DILSLDKKKKQJLILKVKLLHKSNEMLKL 1662

RESULT 6
 ID Q49525 PRELIMINARY; PRT. 1365 AA.
 AC Q49525;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE LMP1.
 GN LMP1.
 OS Mycoplasma hominis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 NC Mycoplasmataceae; Mycoplasma.
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN-MH56;
 RX MEDLINE; 9536982.
 RA JENSEN L.T., LADYFOGED S., BIRKELUND S., CHRISTIANSEN G.,
 RT "Selection of Mycoplasma hominis PG21 deletion mutants by cultivation
 in the presence of monoclonal antibody 552.";
 RL Infect. Immun. 63:3336-3347(1995).
 DR EMBL; U21962; AAA81013.1;
 SQ SEQUENCE 1365 AA; 154982 MW; 23BE4182 CRC32;

Query Match 2.1%; Score 156; DB 2; Length 1365;
 Best Local Similarity 16.8%; Pred. No. 0.57;
 Matches 209; Conservative 187; Mismatches 420; Indels 430; Gaps 52;
 QY 272 TFSAVQTFVQDPLGALYNHDEFLHYQDDWYILSSQIENKPDYIFVYVYGRNDAMD 331
 DB 262 TRNOIQEFINT-----NKNNP--NISE-----LISQLTSKRQ-----SKNSVYDS 319

QY	332	YGGVYITRSPILPESIT--PNLOKA-----AKSYGRFNNPFIITDUSCGEPPLVERLEK	385
Db	320	SNKSDIESANTELOKALANADKVOADNLKSIKEODINNSVANNT-----LSA	369
QY	386	TAEBGEKLLIKREAVIEEVEEVEKVBDETMTLFOLLEGF-----KELOODEENFREL	441
Db	370	KLDDKNTIOQATELEKEVQKADDAITSNNTMASQAKSSLDAKVAELIKKLETE----	426
QY	442	SKEEKEILNELONEATEVERLEFGRALPIRLKLMALAPHSNLANHETIKYVGSKLPGHK	501
Db	426	NKDKAEKFNELKOTRNOIOEF-----	447
QY	502	RFSMGWEDYGSIVAKICTSSRIRIYRPRKSPRICOGLDSDGLOLFEHGHNLSPHAS--I	560
Db	447	-----INTNNPNPYSSELLISQLTSKRDSKNSVTDSSNKSDIESANTEL	489
QY	561	NONVPKNSGCRFPKDVALMWKMGOFAKTVAIFLTSVASKADAVDAKTCOTCLEK	620
Db	490	KOLAKRAN-----ADVQO-----DNLAS	509
QY	621	CRLELACISNPACAAVACLOTCNNRDETECOIKCGDLFENSVDNEFNECAVSRKKCV	680
Db	510	IKEOUNNSVN-----ANTLSAKLTDKDN--TIQOAKTE	541
QY	661	PKSDVCDVFPPDPSPVLVQKFMKDSGKWFITRLANTPDA-----EDCOJHEHHEENK	736
Db	542	LEKE-----YQADDAIKSNNTASMOQAKSSLDAKVAELIKKLETFNKDKKA	588
QY	737	LGVNLMSRIRTPDGGFETRSVAVQKFOYDKRPGIILYNHDEYLLYODWYILSGKVENSP	796
Db	569	KFNELKO-----TRNOIOEFINTNK-----NPN-----YSELISOLTSKRROS--	627
QY	797	EDYIFVYKGRNDAMDQYGGSVLYTRSAVLPESSITP-----EJOQAQKVGROFTFIK	850
Db	627	-----KNSVTDSSNKSDIESANTELOKALKANADKVOADNLAKSIKEODINNSVS	676
QY	851	TDNTCGEPPLPYLELEKKEVEGERITIKKEVELEEVEK-----VRDKEVTL	898
Db	677	NANT-----LSAKLKDKNNTIOQATELEKEIOQANOAIKSNNTASMOQAKSSLD	726
QY	899	SKLPGFEKELD---RDEENFLRELSKEENDVDLGLKMEATEVEKLEFGRALPIRLKLMAVAT	955
Db	727	AKVAELIKKLETFPKDKAEAFNELKOTRNOI-----QEFINTNK-----	766
QY	956	HCFITSPCHDIRFESSDQIGRLGITRRKINGTEFLKILPPIQSD-----LRTGGRS	1009
Db	766	---NNPNSEL-----ISQLTSKRDSKNSVTDSSNKSDIESANTELOKALNTAKAKK	814
QY	1010	S-----RPLSAFSGSGSKGIFDIPL-----PKKNL--KELAPLLKLVG	1049
Db	815	SSIDNELRPL---KNDLOSKIIEEGPIRNTNFSWISSKLETTKKNLAEELT-----	863
QY	1050	VLACAFLLVPASADV-----DALTKTACLLKGCRIELACIANPACAAVACLOTCN	1101
Db	863	-----KADAIKNNPSSSKOALKDSSQOYOKIKNLLETTIIE-----EFGVLETKN	907
QY	1102	NRPDETECOIKGDLFENSVDNEFNECAVSRKKCVPRKSDGLEFPAPD-----PSVLVQ	1155
Db	908	SNIGRLFKRLQAQEPNNSDVKRLKNMEKQFLLSKOKLGNOSTKIDYLTOLSTEMSTQ	967
QY	1156	NNINISD--FNGKWYITSGLNTFPAPCOJHEHTGEDMKVLGNISWRTKILDSOFFIR-	1213
Db	968	ESTIKKIVJNIOAHIRNNLN-----SOYREAD--KLIAN-----MKRGYGVK	1009
QY	1213	--SAVOKFVQODNGGVLYNHDEYLEHYODDW-----YILSKTEKKPEDYIFVYVGRGN	1265
Db	1010	GIESLOKKN--QDLMDSVLSVDS-----LKDDPFKALRLVGVGYTKRPPYSSMFIKKRRS	1064
QY	1266	-DAMDQYGAVVYTRSSVLPSNI-----IPELEKAASIGR-----	1301
Db	1065	IENYQVLRNLLVRENELLIDKAKOLDPRAKETIKFVDENINSIDORARLEKOEILMKN	1124
QY	1301	DRSTPIR-----TDNTCGEPPLVERIKTEYVEGERITVKE-----VEEIEEVEK	1348

```

Db      1125 DLSNFTLNHQKQOFKAKDTPRISLEKLNLEINQYUJLPIIKKAVKSISELKN-KREL 1183
Oy      1349 EKGRTETMLP2HLAGFNLKODENPRELSKEEMEFIDEIKME 1394
Db      1184 EDIIRSNFYLMK-----KVELINKITSELINKOVELKRSNINFE 1220

RESULT      7
ID      090337      PRELIMINARY:      PRT: 1933 AA.
AC      090337;
DT      01-NOV-1996 (TEMBLrel. 01, Created)
DT      01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT      01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE      MYOSIN HEAVY CHAIN.
OS      Cyprinus carpio (Common carp).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC      Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC      Cyprinidae; Cyprininae; Cyprinidae; Cyprininae; Cyprinus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=FAST SKELETAL MUSCLE;
RX      MEDLINE; 97352533.
RA      HIRATAWA Y., WATABE S.;
RA      HIRATAWA Y., WATABE S.;
RT      "Structural differences in the crossbridge head of
RT      temperature-associated myosin subfragment-1 isoforms from carp fast
RT      skeletal muscle."
RL      Eur. J. Biochem. 246:380-387(1997).
RN      [2]
RP      SEQUENCE OF 738-1933 FROM N.A.
RC      TISSUE=FAST SKELETAL MUSCLE;
RX      MEDLINE; 97176447.
RA      IMAI J., HIRAYAMA Y., KIKUCHI K., KAKINUMA M., WATABE S.;
RT      "cDNA cloning of myosin heavy chain isoforms from carp fast skeletal
RT      muscle and their gene expression associated with temperature
RT      acclimation.";
RL      J. Exp. Biol. 200:27-34(1997).
DR      EMBL; D89990; BA422067.1; -.
DR      EMBL; D50474; BA409067.1; -.
DR      PFIAM; PF00612; IQ; 1.
DR      PFIAM; PF00063; myosin_head; 1.
KW      Myosin.
SO      SEQUENCE 1933 AA; 221092 MW; 43DE2593 CRC32;

Query Match      2.1%; Score 156; DB 13; Length 1933;
Best Local Similarity 17.0%; Pred. No. 0.9;
Matches 203; Conservative 181; Mismatches 454; Indels 356; Gaps 42.

Oy      379 LVERLEKTAELCEYLLIKAEVEIEEVEKEVKVDTMTLFFORLEBQFE-----L 430
Db      606 VVQLYOKSALVYALVLVAVPEAAGKGGKGGKGS-----FQYSAVFRLNGLKMTNL 661
Oy      431 OODEENPRELSKEEKE-----LNELOWE-ATEYEKLEFGRALPIR----- 471
Db      662 RSHHGFVRCPLIPNESTPGIMENLVYIHLRONGVVEGIRICTGPFPSRIHYGDEKQRY 721
Oy      471 -KLRLALAPHSNFIANHETIKYVGSGLPGRKFSMGWEDY---GSLVYAKICSSRRIP 526
Db      722 KYLWNSVPIPEGQIDNNKATKEKLGSIDVDHNOYKRGHTKYFFKAGLLGLTEMRDEKLS 781
Oy      527 RYFRKSPRIGCDLDRGJOLFSHGKHNLSFASINONVKGSGCKPFKVDALVWWEKMG 586
Db      782 HLYVTQALARKYVVR--KEFVKKMERRELAISIOYNISFNVAWHP---MMKVYFKIK 836
Oy      587 QFKAIAVAIFILSYASKADAVDAKTCTCLLKECELEDAKICSPACAAVADLOTCNN 646
Db      837 PLKLTAESE---KEMASKEENFEKMKEDLTALAKAKKEELKMWISLVOEKNDLLLOYTSE 893
Oy      647 RPETTCOIKGCDLPENSV-----VEFNECAVSRKK-----CVPKRSDV 686
Db      894 SENLSDAEKRCIGLKSKITOLEGKLTETTERLEDEDEINAEILAKKRRLDEDCSELKKDI 953

```

Qy	687	GDPVYDPSVLQOKEDMDKFSKMWITGGLNPPTDADCOLHEHHEENKVLGNLSWIR	746
Db	954	DDLETLAAVEEK-----HATENK-VKNLTTEMA	982
Qy	747	TPDGFETTSAAVOKFVDPKYPGLYNHNDNEVLYQDWMYIISKVENSPEDIIFYYNG	806
Db	983	SDP-----ESTAKLTKEKKALQEAHQOTLDDLOAEDKYNLTAKATILEQV-----	1033
Qy	807	RNDAMGYSGLYTRS-----AVPESTIP--ELQTAQKVGK----	844
Db	1031	-----DDEGSLSEQKKLRMDLERYKKRLEGGDLTIAQESTIMDENKQSDERIKKKDPE	1085
Qy	844	---DENFIKTDTNCGPE-----PPLVERLEK-----KYEEBERRIILKEVEI	883
Db	1086	ISOFLSKIEDBSLGAOLQKKIKELQIARIEBELIEBELIAERSARAKKOPRADLSRELEI	1145
Qy	884	EEBVEK-----VRDEVTLLFSKLEFG	904
Db	1146	SRLEAGGATMAQIEMNKKREAEFOKMRDLESTLQHEATAAALRRKQADSAVLEIQ	1205
Qy	905	FRELDRDENFLRELSTKEEMVYDGLKMEATYEKLFGRALPRKJMAVATCTSPCHD	964
Db	1206	TNDLRIQKLEKESEKMKEL--DOLSNNEVAKKGN--LEKM--CRT--LED	1253
Qy	965	RIREF--SSDQIGRLGTRIRNGTFLKILPPIQADLRTTGSSK-----P	1012
Db	1254	QISEIKASDENSRLQNMN-----AQRALQIENGEFSQLEEKALVSY	1299
Qy	1013	LSAFSGSKGIFDIVPLPSKNELKELTAPLLKLIVGLVLAACFLVPSADAVDALTKCAC	1072
Db	1300	LTRGQAFQTQOIEDL--KRHYE-----	1321
Qy	1073	ILKGRIFLACIANPACAANVAC--LQICNNRPDELEQIKCG-----DLFE	1118
Db	1321	-----EVAKAKALAAHVAOSARHCDLIRDEYEEQFAKALROGSKANSEVAOMRAKYE	1375
Qy	1119	NSVDFEFCASVSRKKCVPRKSIDLGEFPADPSPVLQVNINISDENGKWTITSGLNPTDPA	1178
Db	1376	TDALORTBELSESKKLLAQRLQDAE-----SIVAVSK---CASLEKTIQR	1419
Qy	1179	FDCOLHEHTGSD--NKLGNISMRILKTDSCGFTTSAAVOKFVQDPNPGVLYNHNEYL	1236
Db	1420	LQSEVEDLMTIDGERANALANLDRKORNF-----KYLAWMKQK--YEESQAEI	1466
Qy	1237	H-YODDWIILSSKFEKNKPEDIIFYYRRGRNDAMDGYGAVVYTRSSVLPNSIIPELEKAA	1295
Db	1467	EAAQKEKRSLSLELFRKNKY-----EELADH-----LETLIREN	1501
Qy	1296	KSIQDFSTIRIDNTGCPALVERIEKTEVGERIIVKEVEETIEEVEKEVEKVGTE	1355
Db	1502	KNLQOEISDSEQLGETGKSIHEIERAKKTVESEKAEIQTALAEFAEGILHEEBSKIIIRVQ	1561
Qy	1356	MTLQO-----PLABEFLNKQDENFVYRELSEKMEFLDEIKK--EASEVER	1400
Db	1562	LELQVQVSEIDRKIAEKDEMEQIKRNSQVLDMSQSTIDSEVRSRNDALRYKK	1615
RESULT	8		
ID	029230		
AC	029230	PRELIMINARY;	PRT; 886 AA.
DT	01-JAN-1998	(Tremblrel, 05, Created)	
DT	01-JAN-1998	(Tremblrel, 05, Last sequence update)	
DT	01-NOV-1998	(Tremblrel, 08, Last annotation update)	
DE	PRIME NTPASE, PUTATIVE.		
GN	AFI032		
OS	Archaeoglobus fulgidus.		
OC	Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;		
FN	Archaeoglobus.		
RP	[1]		
RC	SEQUENCE FROM N.A.		
	STRAIN-VC-16/ DSM 4304 / ATCC 49558.		

RESULT 9
ID 025662 PRELIMINARY; PRT: 1939 AA.
AC 025662;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-AUG-1998 (TREMblrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96V;
RA WERNER E.B., TAYLOR W.R., HOLDER A.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U43145; AAC63403.1;
SQ SEQUENCE 1939 AA; 229001 MW; 5FC4EAC2 CRC32;

Query Match 2.08; Score 151; DB 5; Length 1939;
Best Local Similarity 17.58; Pred. NO. 1.7;
Matches 189; Conservative 167; Mismatches 299; Indels 422; Gaps 50;

QY 380 VERLEKTEEGEKKLIKRAVEIEEVEKEVEKVDTEMTLFORLLEGFKELQOD---EEN 436
DB 95 VKEVEKTEIELEKEL---KLEKEKQINKEVEKELNESEFKROMELLKEKELNINIKEN 151
QY 437 FVRE-----LSKEKEEILNELOME---ATEVEKLFGRALPIRKLMALAPHSNLANHE 487
DB 152 KINNEEITILKREK--LNDIESEVIEKNEKEKL---NYEVNTIKMSL-----D 196
QY 488 TKRYVGSGLPEHKRPSKQWEDYFESIVYVAITCISRRIPRFKRSPPRCGLDGRGLQ 547
DB 197 KLTCEVQEKKNDLEKIN-----KQVIEKENLRELKEKKEKNETIESLDG----- 243
QY 548 SHGKHLSPAHISINONVPRKNGCGKFPKDVALLMVEKKGQFAKTAIYAFILSYASRADA 607
DB 243 -----TIN-----DKNAEKELEI-----SPEEKRM 264
QY 608 VDALKTCGLTCLKECRLELAKCISNPACANVACLOTCNNRDETECOIKCGDLFENSVD 667
DB 265 IEKLD-----SKLIEKEENFAN-----KOAKLE 287
QY 668 ENEECVASKKCVPRKSDVGDVPDPVSLVQKFDKDPGSKMTTGRGLNTPDAFDQOL 727
DB 288 KENELIIEKLIKLESRE-----KDFKSK-----EKFASMEKEL 321
QY 728 HEHTEENKLVGNLS-WRIPTDGGFTTSAYOKFVODPKYGLIYNHNDNEYLLYODDWY 786
DB 322 NTLKSDLSKNACQMEYKLEIKD--LSQSLVEKERE-----IFELIKEL--DD-- 366
QY 787 ILSKVENSPEDYIFVYVGRNDAMDYGGSVLYTRSAVLPSIIPELQTAOKYGRDFN 846
DB 366 -----KINMEKELSSI--ND--KGIDNTYLHSEEEKI-NKLKEKETELNEIHKRYN 413
QY 847 TFIKIDTGTCPPEPILVERLEKKEVEGEPTIIKEVEEIEEVEKVR-----DKFVT 896
DB 414 LELE-----TIKLNKEKELEKKNKAHVEVNTLNLEIK 449
QY 897 LFSKLFEGFKELORDE---ENFLRELSEEMDVIDGLMEATEVEKLEFGALPIRLKMA 952
DB 450 LLEKKTEDAKEGHKNELNELNOLSKLNKEK--DNINKNENTEL----- 491
QY 953 VATHCFSPGHDRIRFSSSDGIGRLGITPRKINGFPLKILPPIOSADLFTTGRSSRP 1012
DB 491 -----NDKISSLNSE-----VN-----IINKKOTLG----- 513
QY 1013 LSAFRSGSKGIFDIYLPKSNELKELTAPLLKLIVGLACAFLLYPSADAVDAKTCAC 1072
DB 513 -----NDIKTLN-----DLINLNKN-- 528

QY 1073 LKNGRIELAKCIANPACANVACLOTCNNRDE-----TECOIKCGDLFENSVD 1123
DB 528 -----EINTSDKNKMKKEDLAMLEMEGKC-----VVID 558
QY 1124 EF-----NECVASKKCVPRK--SDLGEFPADPVSIVONFISPFNKKWYITSGLNTEF 1176
DB 559 EIEKKRYNELFMELEKKEKENYANDND---EISTLNSIYAKD--KEFI----- 604
QY 1177 DAFPCOLHEHTEGDKNKLGVNISMRIKTLDSGFETSAVOK--FVODPNQGVLYNHDN 1233
DB 604 -----EKKEFY--ENKI-----NLFPNFEKKNIENELNSRLKYDNEQ 642
QY 1234 EYLHYODDWYILSKIEKNEKEDYIFVYVGRNDAMDYGGAAYTRSSVLPNSIIPLEK 1293
DB 643 GLIKQIDELNIQKLTKEK--YLOLY-----NDN-----MHMFRSICTKIDMPSSEN 687
QY 1294 AAKSIGRDFSI FI--RTDITGCPPEPALYRIEIKTYEBEGERIIVKEVEIEEVEKEVEK 1350
DB 688 IKGSDLVDEVYAVYIKRDESSSDANDPTTHK--ENVAELEKRAHVALEKHKKEELAK 745
QY 1351 VGRTEMTLFOPLAEGFNE---LKODENFVRELSEKEMEFLDEIKMEASEVEKLG 1403
DB 746 LGEGHKEVVALRGQHKHEETIILIEKHKDVYTKLGQHKENITIKIEEKKDVYTKLG 802

RESULT 10
ID 013246 PRELIMINARY; PRT: 3113 AA.
AC 013246;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE NUCLEAR PHOSPHOPROTEIN MITOSIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 95379848.
RX ZHU X., MANCINI M.A., CHANG K.H., LIU C.Y., CHEN C.F., SHAN B.,
RA JONES D., YANG-FENG T.L., LEE W.H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029(1995).
DR EMBL: U30872; AAA82935.1;
SQ SEQUENCE 3113 AA; 357281 MW; 605779C3 CRC32;

Query Match 2.08; Score 147; DB 4; Length 3113;
Best Local Similarity 18.48; Pred. NO. 5.6; Indels 434; Gaps 60;
Matches 240; Conservative 188; Mismatches 443;

QY 379 IVERLEKTEEGEKKLIKRAVEIEEVE-----KEVE-KVADTEMTLFORLLEGFKELQ 432
DB 946 LSETLSLEKKYSSIIISUNKREIELTQENGTKEINASLNOEKMANLIQK-SESFANYID 1004
QY 433 DEENFVREL--KEKEEL----- 450
DB 1005 EREKSISETSDQKQEKLLILORCEGTGNAYEDLSQKKAQAEKNSKLECLINETSICE 1064
QY 450 -----NELOMATEVEKLFGRALP--IRKLMALAPHSNLANHEITIKYVGSKLPGH-- 501
DB 1065 NRRNELE---QKKEFAKEHOFITKLAFAEERONIMLELEIVQALREMTDNONNS 1120
QY 501 KRFSGWEDYFQSIYVAKTSSRRIPRYFRSPRI-----CCGLDSGLOLFESHGK 551
DB 1121 KSEAGGLQEIWLKKEQKMKQKEVNDLQENQMLKMYKTKHBCQNLSESPIR----- 1175
QY 552 HNSPASHINQVPRKNSCGKFPKDVALLMVEKKGQFAKTAIYAFILSYASKADAVDAL 611
DB 1175 -----NSVAKERSEERNO--CNFRPQMDLEVKEISLDSYNAOIVQLEALRNKEKLOESE 1227
QY 612 KTCTCL-----LKECRLEL-AKCSNP-----ACAN 637

```

Db 1228 KEKCEQHELOTRGDLTSLNLODMOSQSEISGLKDCDEIDAEKYLISGHELTSGNDNAH 1287
Qy 638 VAC-LTOTNNRPDETE--OOIKGDLFENSVDENECVAVSKKCV--PRK--SDVGF- 690
Db 1288 LQCSLOTNNKLNLEKICEILQAEKYE--LVTEIND--SKSECITATRKMAEVEGKL 1342
Qy 690 ----PVPDPSVLQKEDMKDFSGKWFITRGLNP-----FEDAFDC 725
Db 1343 NEVKILNDSOGLHGLVEDIDIGER--GEOPNQHVPYSLAPLDESYSYHLLTSSKEV 1399
Qy 726 QLHEFTENKLVGNLSWRIRIPDGGFTTSAVQKFEVDPKPYGLYNHDEYLLYOD-- 784
Db 1400 QMHFAELQER-----FLSLOSEHKILHDQCO--MSKMSLOLTYDLSL 1441
Qy 784 --DWTLLSKVENSPEDEYFYVYKGRND-----AMDYGGSVLYTRSAVL 826
Db 1442 KAEVLVSTNLNRFQGDVLKEMQJLGEGLVPSLSSCVDPSSSLSSIGDSSEY--BALL 1499
Qy 827 PE--SIIPLOJTAQ--KVRGDFNTFKTDNTGCPPEPL-----VERLEKKEVEGE 873
Db 1500 EOTGDMSTLSNLGAVSANQCVDEVFCSLOEENLTREKTPSPANGVELESLECVYR 1559
Qy 874 RTIIEVEIEEBEVEK--VRDEVTLESKLEEG-----FKELORDEENFLRELISKE 922
Db 1560 OSL--EKLEKMSOGIMKKEIOLEQLLSEBQELDLCKRQYLSENDOMQOKLTSV 1615
Qy 923 EMDVLDGKMEATEVEYKLF-----GRALPIKLAUVATH-----CFTS-- 961
Db 1616 TLEMESKLAEEKQTEQLELEVARLOGLDLSRSRLGIDTEDAIOGRNESCISKE 1675
Qy 961 ----PCHDRITREFSSDGIQRLGITRRKINGTFLKILPPI-----OSADL-- 1003
Db 1676 HTSETTERPKHD-VHQCDKDAQOQDLNLEKITEITGALK--PTGECGEGSPDTNYE 1731
Qy 1003 ----RTTGGS-----SRPLSAFNSGFSKGIQFIVPLPSK----- 1034
Db 1732 PEGEDKTOGSSSECISELSEFSGPNALVPMDFLGNQEDIHNLQLRKETSNNELLHAVIED 1791
Qy 1034 ----NELKELTAPLLIKLVGLACAFILYPSADAVDAKLCACILKGRITLANC 1084
Db 1792 RDRKVESLNNEMKELDSKLQEVOLMT-----KIBAC-----ILEKRI 1830
Qy 1085 IANPACANNAVACLOTCNNRPDETECOIKGDLFENSVDENECVAVSKKCVPRKSDIG- 1144
Db 1831 V-----GEL-----AK--ENSDISE 1843
Qy 1144 --EPPAPDPSVLVONFNISDPNGKWTTSGLNPTFPAFCOLHEFTE--GDNKLYGN 1197
Db 1844 KLEFFSCDHOBELLQREVTSE-----GLNS-----DLEMHADKSSREDIGDNVAKYN 1889
Qy 1198 ISMRKTLDSGFTTSAVQKFEVDPNGVGLYNHDEYLLH-----YODMYLLSKEIENK 1252
Db 1890 DSMKERFLD-----VENELSRINSEKASIEHEALYLEADLEVYOTLEKLCLEKDNEK 1941
Qy 1253 P-----EDYIFVYRRANDAMDYGAVYTRSSVLPNSIIPELKAKAKSGRPFSTPI 1306
Db 1942 OKVIVCLEELSVYTSERNQL--RGELDTMSKITTALDOLSKAKMEKIOELSESQSECL 1398
Qy 1307 RTDNTGCPPEALVBRITKYVEBGERIIVVEVEIEEB--VEKEVEKVGRTEMTLFORIA 1363
Db 1999 HCIOVAEAVERKTELLQTLSSDVSELKDKTKHLOKLSLEKDSQALSTKCELENQIA 2056
Qy 1364 EGENELKODEENFYRE-----LSKREMEFLDEIK--MASSEVEK 1400
Db 2059 ----QLNKEKELLYKESSESLOARLSESDEYKLNVSALAEALVER 2099

```

```

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DE 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
GN NDM1P.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota, Fungi, Ascomycota, Hemiascomycetes, Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA MURPHY L., RICHARDS C., HARRIS D.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L0001287; NDM1.
DR SGD; L0001287; NDM1.
DR PFAM; PF00169; PH; 1.
SQ SEQUENCE 2748 AA; 313030 MW; FEEETEDD CRC32.

Query Match 1.9%; Score 146; DB 3; Length 2748;
Best Local Similarity 17.9%; Pred. No. 5.4;
Matches 283; Conservative 232; Mismatches 581; Indels 486; Gaps 72;

Qy 62 SSFSDSHCKDKSQICSDISFEELQRFDLKGMILILEKORQF-----IQALVIVLC 115
Db 748 SAYEDLVKCKENDV-----EFLKEKSAKLGHTVVSSEYSELQRYSELEKEVEQPS 800
Qy 116 TFIIVPRVAVDAKLTQACLLKCEKRIELAKCIANPSC-----AANVACLOTCNNRPDET 169
Db 801 LAYIVHAHAKATDH-----HILSDSAYELVKCKENPDMERLKSASAKLGHTVVSNEAYSEL 856
Qy 170 ECGI-----KCGD--LFENSVDQFNECAVSKKCVPRKSDVGEPPVDRNAV 215
Db 857 EKKLEQPSLAVIYEHAKARDHILSDSAYEDLVK-----KENS DV--EF----- 900
Qy 216 VQNFNMKDFSGKRYITSGINPTFPAFCOLHEFHEHENDKIVGNLTWRKITLDGGEFTTSA 275
Db 900 ----LKEKSAKLGHTVVSNEAYSELEKLEQ-----PSLAVIYEHAKARDHILSDSA 948
Qy 276 VQTFVODPDLPGLYNHDEYLLHQQ-----DDMYLLSQIENKPDQYFYVYRG 324
Db 949 YE-----DLVKCKENPDMERLKSASAKLGHTVVSNEAYSELEKLEQPSLIVYEHAKA 1002
Qy 325 RND--AMQGYGSV-----IYTRSPILPESITPN-----LQRAKSVGRDF-- 364
Db 1003 TNNHLLSDSAYEDLVKCKENPDMERLKSASAKLGHTVVSNEAYSELEKLEQPSLEYLVE 1062
Qy 364 ----NNFITDTS-----CGPEPLVBRLEKTAEBEKKILINEAV--EIEVEYEKE-- 409
Db 1063 HAKATNNHLLSASAYELVKCKENDVDFLEKESAKLGHTVVSNEAYSELEKLEQPSLE 1122
Qy 409 --VEKYRDTMTLFFORLLEGEKELQODEENFVRELSEKEKILNE--LOMEA--TEVEKLF 463
Db 1123 YLVEHAKATNNHLLSD--SAYELVKCKENDVDFLEKESAKLGHTVVSNEAYSELEK- 1179
Qy 464 GRALPRKRLMALPHASNFANHETIKYVSGSKLPGHKRFSGMEDYFGSIYAVK----- 519
Db 1179 ----KLEQ--PSLAVIYEH-----AKATDHLLS--DSAYEDLVKCKENPDV 1217
Qy 519 ----ICSSRRIPRYFRKSPRICGDSNGL--QPSHGKHNLSPPHASNQNV 564
Db 1218 EFLKESAKLGHTVVSNEAYSELEK-----LEQPSLAVIYEHAK--ATDHLLSDSA 1268
Qy 565 PKGNSGCKFPKQDVALMWKQWOPAKTAIYAFILTSVASKAD-----AVDAKLTCTCL 617
Db 1269 YEDLVKCKENPDMERLK--EKSALGHTVVSNEAYSELEKLEQPSLEYIVYEHAKATNNHLL 1327
Qy 618 LKECHLE-LACISNPAC-----AANVACLOTCNNRPDETECOIKGDLFENSVDDEFN 670

```

```

Db 1328 LSDAYIEDLVCKENPDMFLKESAKIGHTVWSNKEYSELEKLEQPSL----- 1378
Qy 671 ECASVRKKCKVPRK-SDVGF-----PVPDPSVLVQKFDKMGKSGKWFTRGLNPTFPAF 723
Db 1378 ELYLVHAQIOGKISIDSDNTLANPMSMEDWASKLOKLEVOYVSNDEYI--ALKNTMEKP 1435
Qy 724 DCG-----LHEFH-----TEENKLVNLSMRIRTDGGEFTTSAQKFPVOD-PKYPGLLYN 773
Db 1436 DVELLRSLKGYHIIDTTYVELVSNFN-----SPLKFTIEKASKGYRLI 1482
Qy 774 HONEYL-----LYODMYLLSK-----VENSPE-----D 798
Db 1483 ERENEHIDLNRITTSKEIDNFCQICGALDSKEYERLKNLSNPKFEENALLD 1542
Qy 799 YTFV---YKGRNDAMDGVGSVLYTRSAVLPESTIIPLOTAOKVGRDFNTFKTNTNC 855
Db 1543 LVLVDKTEYQAMKD-----NASNKKSLIPSTKAL-----DFVTM----- 1577
Qy 856 GEPPPLVELEKKEVEGERTIIKEVEEIEEVEKVRDKEVTLFSKLEGEKELORD--- 912
Db 1577 -PAPOLASAKSSLOK-----RTLSDIENELKAL-----GYAIREKNLNP 1616
Qy 912 -EENFLRELKKEEM-----DYDGLKMEATEVEKLEFG-----PALPIR 948
Db 1617 LKRPYDNASKNDVNLKSGKSLVPLSTEEYDNMRKHTKLINLIGDPSIDFLKECEKY 1676
Qy 949 KLMVAATH-----CFISPCDRIREFSSDGIQRLGTRKINGTFLKILPPIQAD 1001
Db 1677 QMLIKHDEYKOEALINPGEYFLEKASALGY-----ELSEVELDMKQOMIDSPD 1729
Qy 1002 LRTGGRSSRPISAFSGSGKGFIDVPLPSKNELEKETAPLLKLVGLACAPLIYPSA 1061
Db 1730 IDVMOKAARN-----EMVLL--RNEKE-----ALQKKIEYPSLTFIEKXA 1770
Qy 1062 -----DAVDALKTACLLKCGRIELAKCIANPACANVACLOTCNNRPDETEQICGC 1114
Db 1771 GNMKILVDQLEYDET-----IRRC-----NHPTMELEESCH 1802
Qy 1115 DFEENSVDDEFNCAVSRKKCVPRK-SDLGEPFAPDPSVLYQNFINSIDNGKWYITSLN 1173
Db 1803 HL--NLVLLDQNEYSTLRELENNRVEDLINTLSKNTLYAIPNTTYOGLIGYE-----N 1855
Qy 1174 PTFDAFCOLHEFHTEGDGNKLVGNISMRITKIDSGEFTTSAAVQKEVODPNQGVLY--- 1230
Db 1856 PNDYLYKDSLNK-----MDYVALISQDYLWAKYKQQLDYLIKIS 1896
Qy 1230 -----NHONEYLHODDWTYLLSSKIENKPEYIIFYYGRDANDG 1270
Db 1897 SEKIDHIVPLSEYNLMATVYRNPSLYLKEKAVLNHILIKEDDKNLL-----AVSE 1950
Qy 1271 YGGAVVYTRSSVYLNSTIIPLEKAASIGRDPSTFTRTNTGCPALVERIEKTEVEEGE 1330
Db 1951 HPYVHUSSEKASLNKLVLYKD-----DFATMSRSIE---KTIIDPLSTKALSMG- 1998
Qy 1331 RIIVKEV-----EEIEEVEKEVEKVRTMPLFORLAEFENELKODEBENFRELKSEEM 1385
Db 1998 KIIVNEGTHKRNKELISEPSEFLIMAKAQGLIISEKYESLRQIDRPSLDVLYKERA 2057
Qy 1386 EFLDEIKMEASEVKLEFGKALP 1407
Db 2058 AIFDSIIVENIEYOQLVNTTSP 2079

```

RESULT 12

Q26024 PRELIMINARY: PRT: 839 AA.

AD Q26024: 01-NOV-1996 (TIREMBLrel. 01, Created)

DT 01-NOV-1996 (TIREMBLrel. 01, Last sequence update)

DE TOCISOMERASE I.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96001260
RT TOSH K.T., KILBERY B.J.K.:
RT "The gene encoding topoisomerase I from the human malaria parasite
RT Plasmodium falciparum."
RT Gene 163:151-154(1995).
DR EMBL; X83758; CA58716.1; -.
DR PFAM; PF01028; Topoisomerase_I; 1.
KW Isomerase; Topoisomerase.
SQ SEQUENCE 839 AA; 98109 MW; 1CAFA3F0 CRC32;

Query Match
1.9%: Score 144.5; DB 5; Length 839;
Best local similarity 17.3%; Pred. No. 1.3;
Matches 191; Conservative 176; Mismatches 390; Indels 347; Gaps 46;

Qy 352 LOKAAVGRPFNFTTDSNGPEPLVERLEKTAEGEKLKKEAVEL--EEVEKEV 409
Db 24 INKIKONIG--NN--KSCNSRSSKESIKKOKNSDELGIKKNTKSLGIKKEEKKKOI 78
Qy 410 EKVDRTEMTLPRLIEGFEKLODEENFRELKSEKEIINELOMEATEVEKLEFGALPI 469
Db 79 SKRKSNELEKRNKLEKGGKATVEKSRIV---KDTKLINVIKKEQONKK-----EK 128
Qy 470 RKLPMALPHSNFLANHETIYYVSGKLPGHKRPSGWDYFGSIYVAKICSSRLRPIF 529
Db 129 KLLKKS---EENF---EPINRW--WEKIDQTDIQWNY----- 159
Qy 530 KRSPLRCGLBGRQLGFSHGKHNLSPAHSINOVVPGNSGCKPRKVALMWEKMGQFA 589
Db 159 -----LEHNGI-IF-----SPDY-VQHHVP-----IYKSIKTEL--NA 188
Qy 590 KTAIVAFILTSVASKADAVDAKLTCTCLKECRLELAKCISNPACANVACLOTCNNRPD 649
Db 189 KSEELATVYCS-----AIGSDYCTKEKEFILNFTFTIN-----SLENDNITIKQ 231
Qy 650 ETEQCIKCGDLFENSVDDEFNCAVSRKKCVPRKSDNGDFVPPPSVLVQKFDM----- 704
Db 232 ENETRLKKGDIENKEFID-----FMPKIDHLKREKLNKTKEE 271
Qy 704 KDSGSGWITTRSLNPTFPAFCOLHEFHTEENKLVNLSMRIRTPDG-----FTTSAAV 758
Db 272 KEEKKRMKEKELPYAYALVDIRKISSNKAEPGLFRGGEHPKQGLKRRIFPDDV 331
Qy 759 QKFPDPPKYPGLYV-----HDNE--YLYVODDWTYLLSKVENSPEYIIFYYVKG 806
Db 332 INISKAPYPRLYDNGCGHNGMDIHDNKVTWLAYKD-----SINDQIKYFELS 381
Qy 807 RNDAMDGVGSVLYTRSAVLPESTIIPLOTAQKVGDFNTFIKT---DNTCGPEPPV 862
Db 382 AQSKEFGYKDKLMYENAR-----KLKSCVHKIREYKNNKKNKIIDQDGTAVYLI 433
Qy 863 ERLEKKEVEGERTIIKEVEEIEEVEKVRDKEVTLFSKLEGEKLODEENFRELKSE 922
Db 434 DFLALRV--GGER-----DIDEADTV-----GCCSLRVEHLSFADHDPK 472
Qy 923 EMDVLDGLKMEATEVEKLEFGALPIRKLMAVATHGFTS---PCHDRIFESSDGIQRLG 979
Db 473 SVD-----SKQQLNDEKVKIPLPNTLESISEDCYITLDELGKDSIRYFWT----- 521
Qy 980 ITRKRINGTFLKILPPIOSADLRTTGGRSSRPISAFSGSGKGFIDVPLPSKNE-LKE 1038
Db 521 ---VKIDKQAVINII-----IFCKNKNRDEGVFDQITGSKLEIYKE 559
Qy 1039 LTPALLKLVGVACAFLLIYPSADAVDAL--KTCACLLKGRILELAKCIANPACANVAC 1096
Db 560 IMPLTSKAVRPTNASTITDQOLKRIKEYYGGTYSVSG--ETELHK-----SK 607
Qy 1097 LQTCNNRPDETCQIKGDLFENSVDDEFNCAVSRK---KCYRKSDLGEPFAPDPSY 1152
Db 608 KRKSSHLTSDTNILSDASDSTINDVNNVEYDENGINKLSYATTVGKENDVDKNSP-IEV 666

```

QY 1153 LYONENISDENGKMYIISGLNPTFDACDQLEHFEHTEGDNKLVGNISWIKITLDSGFTTR 1212
 DB 667 DVSNIN-----ELINFNANNREYALICNHNORSIPKCHDTM 703
 QY 1213 SAOKEVQADENQGVLYNHD-NEYLYODDWMYILSSKIKENKPEDEYIFVYRGRNDAMGY 1271
 DB 704 SKIKQJE-----LYNEDIKEYKKY-----LOHLKNSDKKITFV----- 739
 QY 1272 GGAIVYTRSSVLNPSILPELEKAKSIGDPSFTFRTDNTGCPDPALVERIEKTEVEEGER 1331
 DB 739 -----SKYSTLDGTLRP-----NKVKENMK 758
 QY 1332 IIVKEVEEIEEVEKEVEKVRTEMTLFORLAEGNELKODENFVRLSEKMEHFLDE- 1391
 DB 759 -----EESCKKLL-----ITLIKVELLNQMKVDRDNTKIALGTSKINYMOPR 802
 QY 1391 -----IKMEASEVEKLFGLKALPIR 1409
 DB 803 ITVAFCKKEEIPLEKVFENRSLRLK 826
 RESULT 13
 OS1465 PRELIMINARY; PRT: 2166 AA.
 AC OS1465;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DE 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 GN BH0512
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 35210 / B31.
 RX MEDLINE: 98065943.
 RA FRASER C.M., CASTENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
 RA LATIGRA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
 RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
 RA PETERSON J., KERLAVAGE A.R., QUACKENBUSH J., SALZBERG S., HANSON M.,
 RA VAN VUOT T., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
 RA UTERBACK T., MATTHEY L., MCDONALD L., ARIACH P., BOWMAN C.,
 RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
 RA SMITH H.O., VENTER J.C.,
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi."
 RL Nature 390:580-586(1997).
 DR EMBL: AE001153; AAC6876.1; .
 DR TIGR: BH0512; .
 KW Hypothetical protein.
 SO SEQUENCE 2166 AA; 254243 MW; F2899A0A CRC32;

Query Match 1.9%; Score 144; DB 2; Length 2166;
 Best Local Similarity 16.8%; Pred. No. 5.1;
 Matches 268; Conservative 275; Mismatches 556; Indels 500; Gaps 71;

QY 279 FVQDDPLPGALYNHNEFLH-----YQDDWYILSSQIENKPDYIFVYRGRNDAMGYCG 334
 DB 641 -----SQIKTKEEFLSLIQOKGIGLSSVNDSDHT-----OKRAIDHGS 666
 QY 335 ---SVIYTRSGFL-----PESITPLOKAAKSVGRDENFTTDSGCPPLVERLE-- 385
 DB 687 WKDEILALNKSLLIDIKVSESELLSSATLKIESLEKDVN-----DAMEYV 730
 QY 385 --KTAEEGKLLIKFAV-----IEEVEKEVEKVRDTEMTLFORLLEGF 427
 DB 731 LAKTGD I-ESLYIEKYKELKDMYSQSDEALIGKEFINQTEGIIKDKSVFMEDLNKKE 789
 QY 428 KELQDDEENFV-----RELSEKEEILNELMEATE-VEKLFGRALPIRLKRA 475
 DB 790 -----DDKNMFVYSKIECDYKLDKFKIESDILNFRSDLINEIES-----KLOIV 836
 QY 476 LAPHS-----NELANHETIKYVGSKLPGHKRFSQWEDYFGSIIV--AK 518
 DB 837 SNIKSDNOKQIDDFLRISKDILNRKDSINNEVDSKLS-----DMQSKLINEITVKIEN 889
 QY 519 ICSSRRIP-----RYFRKSPRICCGIDSRGLQFLSHGK----- 552
 DB 890 ILSGKVDLDLIDSEVYTKIKELKFIES-----LESTYLEKIDFRRQGAISDELLQ 943
 QY 552 ----HNSPASHINQVYKPG-----NSGCKFPKDVALLMWEKVGOPAK-TAIVAFILISV 601
 DB 944 DIMNHFNKRELEENLSKFFAVALNSEEVEKREVSLLQDKRTDIASFQAMIDITLSTL 1003
 QY 602 ASKADAVDALCTCTCLKEORRELAKCISN-PACANAVAC----- 641
 DB 1004 NKFENDIN-----KEINGKINEYISNRYGYSNISSKLENIEMHEINLSRLADR 1054
 QY 641 -----LQTCNNRPDETECOI-----KCGDLFE-----NSVDEFECAVSR-K 677
 DB 1055 IDSLSKGDENILQKLESFVSQYQVEKPELKYKDLTDGEAKINKVLKEIQYKSRLE 1114
 QY 678 KCVPRKSDVGFPPVDPSSVLYQKFDKDKDSGKFFITRGILNPTFDACDQLEHFEHTEENKL 737
 DB 1115 EADYRRTI-----DNDIN-----QAKERGE--TINELKNNIESKSEPLDLYKERKYL 1162
 QY 738 V-GNLSRRIRTPDGGFTTRS--AVQKEVQDPKYPGLYHNDNEVLYQDDWYILSSKVEN 794
 DB 1163 IESNEEERYST--FLISEGALISK-INDELIKTLITSDEN-----LOIKISE 1206
 QY 795 SPEDYIIVYKGN-----DAMDYGGSVLYTRSAPVPSITPELOTAQKV 841
 DB 1207 MDQNFELIEORSKDLIEFEKELQDKIKDCY-GFINSQFEIKAGVEENIKNHFVCIKV 1265
 QY 842 GRDFNPFIKTDNCGEPPLVERLE--KKVEEGERTITKEV-EETEEVEKVRD-----KE 894
 DB 1266 ----NTLI-DDDLVYKENEIHKRIDSLKISTESTPISIEKNLNDKVGSGCIDYADENIKY 1320
 QY 895 VTLFSKIFEGFKLQDDEENFLELSKEEMDVLDGIKMEATEVEKLFGRALPIRKIMAVA 954
 DB 1321 IELEERENQQLNLENKIKIDNALIKADINLALSOYDGLERKAYDMYDEFSERL--NSYATL 1378
 QY 955 THCFSPCHDIRIFPSSDQIGRLGITRRKING-----TLFLKIILPIQ 998
 DB 1379 SEEFKSNKEMT--FLEBQLNKLKULESDLNVERKVIKESYHNVSHLLEDEF 1436
 QY 999 SADLRTTGRSSAPLSAFRSFGSKIFDIYLPUSKNELEKETAFLIKLVGLVACALIV 1058
 DB 1437 FKDLKIRGEHLKYSLENFIASYNDKIONLEVDLSN----- 1473
 QY 1059 PSADAVDALCTACILKGRILACIANPACANAVACLOTCNNRPDETECOIGDLFE 1118
 DB 1473 -----LENKTELQSRDLDEQKMDKDENFYLDPTKEFSSKKRDMOSEIA--LME 1521
 QY 1119 NSV--VDEFNECAVSRKCVPR-----KSDIGEPAPDPVVLQNNNISDF-----NG 1164
 DB 1522 TNLTKGVDEDFVFNKQSIIDSWSFLNKDKVKQWQEKSYSTIERKRLNLGLGKSPEND 1581
 QY 1165 KWIITISGLNPTFDACDQLEHFEHTEGDN-----KLVGNISWIKITLDSGFTT 1211

RA TUMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
 RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
 RA TRUST T.J.,
 RT "Genomic Sequence Comparison of Two Unrelated Isolates of the Human
 RT Gastric Pathogen Helicobacter pylori."
 RL Nature 397:176-180(1999).
 DR EMBL; AE001481; AAD06047.1;
 SQ SEQUENCE 1819 AA; 207505 MW; 8809P345 CRC32;

Query Match 1.98; Score 142.5; DB 2; Length 1819;
 Best Local Similarity 19.0%; Pred. No. 4.9;
 Matches 267; Conservative 199; Mismatches 524; Indels 417; Gaps 68;

QY 178 LFENSVVDQFNECAVRKCKVPRKSDVGFPPVDRNAVQNNMKDFSGKWITTSGLNPT 237
 DB 244 LERSRSLFHYF-----VPLEDKSSRFS-KDRNLVYND----- 274
 QY 238 FDAFDQQLHEFHENDKLIVGNLTWRIKTLDGCFTRSAVQTFVODPDLPGALYNHNEFL 297
 DB 274 ----EIQIRQ----EYNRLKERNEKGNMIDKNLF-----ENDDPN-RTIYN----- 312
 QY 298 HYODNVISSQIENKPDYIEVYGRNDAMDYGSGVYTRSPITLPEIIPNLOKAK 357
 DB 312 ----YLNIAELIEDK--NPLRAFYECISN-----GGN--YEELKLIKID--KILQDQWK 354
 QY 358 SVGRDNFNFTTNSGCPPELVERLEKTAEEGR--LLIKEAVEL-----EEV 405
 DB 355 KILEAVINDCI--KNKATBEERIKCDLIKDEKLSLNQOKVOYALDCLKNATDEER 411
 QY 406 EKEVEKVRDTEM-TLFORLLEGEFKELODENFVRELSEEEK-EIINELOMEATEVEKLF 463
 DB 412 KECILKINPEIEKERKRELQLOEKYKDCIKNAKTAEKNECKIGLSKRA--IERLK 469
 QY 464 GRALPIRLKLMALAPSNLANHETIKYVGSKLPGHKRFSGWEDYRGSIVANICSSR 523
 DB 470 QQALDCLKNAKTDEERKECLKN--IPODLOKELLADM-----SVAKYKDCVSR 515
 QY 524 RIRPRYFRK-----SPRICGSLDSRGLOLFSHGKHLSPAHNSINONVPK-----GNSG 570
 DB 516 ARERKQCEKILLPEAKKLLBNQALDCLKNAKTDEERKECL-KNLPDLODILAKES 574
 QY 571 CFFPKDVA-----LMWEKMGOFAKTAIVAIFLTVASAKADAVDALKTCTC 616
 DB 575 LKAYKDCASQATKAEKCEKILPE-----AKKLEEAKEKSVKAYLDCVSAKT--EA 628
 QY 617 LKEC-----KLEIAKCI SNPACANACLOTCNNRPDETBCQ---IKGDLPE 662
 DB 629 ERKECEKILLPEAKKLEBAK---KSVRAYLDCVSKAKNEAERKECEKILLPEAKKLE 684
 QY 663 NSVYDFNECAVSRKCKVPRKSDVGFPPVDSVLQKFDMDKFGKMFITRGLNPTEDA 722
 DB 685 NQALD-----CLNNAKTDEERKECLDLP-----KDLOKVLAKESAVYILDC 727
 QY 723 FDCQLEHFTENKLVGNLSWRIPTPDGFFTRSAVOKFVODPKYPIIY----- 773
 DB 728 VSKAKNEAERKECE-----KILTPPE-----ARKLLEAKKSVKAYKDCVILARNE 772
 QY 773 --NHUNEYLLYODDWIILSSKYENSPEDYIYVYKGRNDAMDYGSGVLYTRSAVLPEST 830
 DB 773 KKKQCEKILLPEAKKLEES--KRSVAYLDCVSKAKNEAERKECEKILLPEAKKLEEA 831
 QY 831 IPELOTAOKVGRDENTFIKIDNTGCPPEPLVERLEKKEVEGERP-----IIEVEEI 883
 DB 832 KRSVAYKDCVSRARNE--KEKQEC--EKLLIPEAKKLEESKSVKAYLDCVSKAKNEA 887
 QY 884 E-EVEKVDKREVTLESKLFESEKEL-----ORDENLRLSEKNE-----EM 924
 DB 888 ERKECEKILLPEA---RKLLLEAKESVAKYKDCVSRARNEKQCEKILLPEAKKLEEN 944
 QY 925 DVLDDGKIMATEVE--KLFRALP---IKLMAVAITHCTSPCHDRIRFPSSDDGIGRLGI 980
 DB 945 QALDCLKNAKTAEKRCVCKDLPKDIQKVLV-----KESVRYLDCVSKAKNEA 994

QY 981 TRKRNITFLIKLIPPIQASADLTFTGGRSRPLSARSPGSGIPIVLPSPKNEIKELT 1040
 DB 995 ERKECE-----KILTP-----EARKLLEAKESVAKYKDCVSRARNE-----KEKQCEKILL 1041
 QY 1041 APILLKLV-----GVLA-----CAFLVPSA-----DAVDALKTC-- 1071
 DB 1042 TPEARKLLEQEVKSKVAYLDCVSRARNEKEKQCEKILLPEAKKLEENQALDCLKNAT 1101
 QY 1071 ACLKGRLEL-----AKCIANPACANACLOTCNNRPDETBCQ---IKGDLFENS- 1121
 DB 1102 EAEKRCVCKDLPDLOKVLAKESVAYLDCVSRARNEKQCEKILLPEAKKLEESK 1161
 QY 1121 -VDFNECAVSRKCKVPRKSDVGFPPVDRNAVQNNMKDFSGKWITTSGLNPTDAF 1179
 DB 1162 KSVAYLDC--VSKAKNEAERKECEKILLPEAKKLEE-----AKESVAY 1205
 QY 1180 -DCQLHEFHTEGNKLVGNLSWRIKTLDSGFFTRSAVOKFVODPDPNPGVLYNHNDNYLHY 1238
 DB 1206 KDC-----VSRARNEKQ-----ECKILLT 1226
 QY 1239 QDDWYILSSKTEKPKPDYIEVYGRNDAMDYGAGVYTRSSVLPNSIIPPE-----LEKA 1294
 DB 1227 PEARKLLEQEVKSKVAYLDCVSRARNE-----KEKQCEKILLPEAKKLEEK 1275
 QY 1295 AKSIGRPFSTFITDNTGCPPEPLVERLEKTAEEGRILYKEVEELIEVEEVEKVGRT 1354
 DB 1276 RQOKRAIKDCLK--NADPNDRALIMKCLDGLSDEEKL--KYLQEARERKAVLDCLKTART 1331
 QY 1355 E-----MTLFORLAGF--NELKODENFVRELSEEEK-----MEFLDE----- 1391
 DB 1332 DEERKQNNIYSDLIQEIQKKAQNNQNN--QSKTERLHQASECIDINDPTQDAIHO 1388
 QY 1391 -----IMKASEVEKLF 1402
 DB 1389 CLEGLSDSERALILIGIKQADEVDRIY 1415

Search completed: October 13, 1999, 22:45:04
 Job time: 722 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 1999, 04:02:19 ; Search time 2137.21 Seconds

(without alignments)
2319.369 Million cell updates/sec

Title: US-09-075-375-5

Perfect score: 1555
Sequence: 1 ccaagcgtccggttgggtg.....gataaaaaaaaaaaaaa 1555

Scoring table: IDENTITY_NUC

Searched: 679419 segs, 1590154680 residues

Database : GenEmbl.*

1: gb_dal.*
2: gb_da2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sts.*
15: gb_sy.*
16: gb_un.*
17: gb_vl.*
18: em_fun.*
19: em_htg.*
20: em_hum1.*
21: em_hum2.*
22: em_in.*
23: em_om.*
24: em_or.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pl.*
29: em_ro.*
30: em_sts.*
31: em_sy.*
32: em_un.*
33: em_vl.*
34: gb_htg1.*
35: gb_htg2.*
36: gb_in1.*
37: gb_in2.*
38: gb_dal.*
39: em_da2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1555	100.0	1555	8	ATU44133 Arabidopsis

2	958	61.6	13840	8	AC003981	Complete
3	922	59.3	2884	8	ATU39452	Arabidopsis
4	624.2	40.1	1589	8	NTU34817	Nicotiana
5	611.8	39.3	1760	8	LSU31462	Lactuca sat
6	330.4	21.2	89493	8	AC006932	Genomic S
7	92.4	5.9	7218	8	166494	Sequence 14
8	54.8	3.5	2813	7	CHOOOREFS	X64618 O.odorata c
9	54.2	3.5	100000	9	AP000067	AP000067 Homo sapi
10	53.2	3.4	200000	34	AC004670	AC004670 Homo sapi
11	52.8	3.4	169931	11	AC005822	AC005822 Homo sapi
12	52.8	3.4	153053	34	HSJ9612	AT009612 Homo sapi
13	52	3.3	4238	12	RNCN641	AT000496 Rattus no
14	51.8	3.3	1228	3	MRU59897	U59897 Macropus ro
15	50.6	3.3	3161	12	MUSCDP	L37092 Mus musculu
16	50	3.2	100000	9	AP000066	AP000066 Homo sapi
17	49.8	3.2	178108	35	AC006151	AC006151 Homo sapi
18	49.6	3.2	160687	11	AC004957	AC004957 Homo sapi
19	49.6	3.2	143749	11	AC004962	AC004962 Homo sapi
20	49.2	3.2	17522	12	MUSMHC4H2S	M17440 Mouse MHC (
21	49	3.2	118595	9	HS288L1	Z82196 Human DNA s
22	48.8	3.1	39229	11	AC004755	AC004755 Homo sapi
23	48.8	3.1	323792	34	AC004965	AC004965 Homo sapi
24	48.8	3.1	207984	35	AC006329	AC006329 Homo sapi
25	48.6	3.1	153147	9	HS13D10	AL021407 Homo sapi
26	48.6	3.1	196361	11	AC005386	AC005386 Homo sapi
27	48.6	3.1	149886	12	NMHC438N12	AF049850 Mus muscu
28	48.6	3.1	203407	34	AC006174	AC006174 Homo sapi
29	48.6	3.1	114631	42	AC006010	AC006010 Homo sapi
30	48.2	3.1	2117	3	CFU46916	U46916 Canis famli
31	48.2	3.1	13095	10	HSFGR4G	Y13901 Homo sapien
32	48.2	3.1	134450	11	HUAC003049	AC003049 Homo sapi
33	48.2	3.1	142228	34	AC004123	AC004123 Homo sapi
34	47.8	3.1	124191	11	AC005153	AC005153 Homo sapi
35	47.8	3.1	162691	12	AC003061	AC003061 Mouse Chr
36	47.8	3.1	108752	12	AF084363	AF084363 Mus muscu
37	47.6	3.1	148382	42	AC006101	AC006101 Homo sapi
38	47.4	3.0	665	14	HDMU71202	L30455 Human STR u
39	47.4	3.0	180385	42	AC007461	AC007461 Homo sapi
40	46.6	3.0	92657	8	ATT15N24	AL078465 Arabidops
41	46.4	3.0	3811	11	HSU88153	U88153 Homo sapien
42	46.4	3.0	3121	11	HSU88154	U88154 Homo sapien
43	46.2	3.0	2540	7	OBCHORS	X64615 O.berterian
44	46.2	3.0	161475	9	HS164C20	AL009029 Homo sapi
45	46.2	3.0	64045	34	PFMAL4P1_3	Continuation (4 of

ALIGNMENTS

RESULT 1
ATU44133 1555 bp mRNA PLN 15-JUN-1998
LOCUS Arabidopsis thaliana violaxanthin de-epoxidase precursor (AYDE1)
DEFINITION MRNA, complete cds.
ACCESSION U44133
NID g1465734
VERSION U44133.1 GI:1465734
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 1555)
AUTHORS Bugos,R.C., Hieber,A.D. and Yamamoto,H.Y.
TITLE Xanthophyll cycle enzymes are members of the lipocalin family, the
first identified from plants
J. Biol. Chem. 273 (25), 15321-15324 (1998)
MEDLINE 98288253
REFERENCE 2 (bases 1 to 1555)
AUTHORS Bugos,R.C. and Yamamoto,H.Y.
TITLE Direct Submission

JOURNAL Submitted (02-JAN-1996) Robert C. Bugos, Plant Molecular
Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI
96822, USA

FEATURES

source

Location/Qualifiers
1..1555

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

1..1555

/gene="AVDE1"

45..1433

/gene="AVDE1"

/codon_start=1

/product="violaxanthin de-epoxidase precursor"

/protein_id="AAC50032.1"

/db_xref="PIR:1465735"

/db_xref="GI:1465735"

/translation="MAVAHCFSPCHDRIRFSSDDGIGRLGTRKINGFLKIL
PIQSADLRTRGRSSRLSAFSGSKGFDIVLPSSKEIKETAPILIKIVGLA
CAFLIVPADAVDALTKACILKGRICRIELKCIANPACANVACLOTNNRDETECO
IKGDIFENSVDPEFNCAYSRKCVPRKSDLEFPAPDSVLVONNISDFNGKWTI
TSGLNPTFDADCOLHEFHTEBGDKLVNLSWRTKLDSCFFTRSAVOKFVODPNPG
VIVHDEYLAHYODWYILSKLENKPEYIFVYRGNDAMDGYGAAVYTRSSVLEP
NSIILELEKAKRSIGRDFSTFIPTDNTGCEPALVERIEKTEVEGERIIIVKEVELEP
EVEKEVEKVRTEMTLFLORLAEGFNEIKODEENVRFLSKEMEFIDEIKMEASVEKE
LFGKALPIRKVR"

transit_peptide 45..383

/gene="AVDE1"

mat_peptide 384..1430

/gene="AVDE1"

/product="violaxanthin de-epoxidase"

BASE COUNT 485 a 285 c 376 g 409 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 1555; DB 8; Length 1555;
Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ccacgcgtccgcgtggtggtggaagattagatgctgaagaatgacgtatgctcac
Db 1 ccacgcgtccgcgtggtggtggaagattagatgctgaagaatgacgtatgctcac
QY 61 attgttcaactcaactgtgcatgacgtatgcttcttccaaagtatgataatg
Db 61 attgttcaactcaactgtgcatgacgtatgcttcttccaaagtatgataatg
QY 121 gtagcttgatgcatgacgaagaagaagatgacgtatgcttcttccaaagtatg
Db 121 gtagcttgatgcatgacgaagaagaagatgacgtatgcttcttccaaagtatg
QY 181 caatcaagaatgctgcatgacgaagaagaatgctgctgacgtatgcttcttcc
Db 181 caatcaagaatgctgcatgacgaagaagaatgctgctgacgtatgcttcttcc
QY 241 tcaggtgacgattcttcaaggagatattgacattgtgcattacacaaagaatg
Db 241 tcaggtgacgattcttcaaggagatattgacattgtgcattacacaaagaatg
QY 301 tgaagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
Db 301 tgaagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
QY 361 ttattgttccatctgacatgacgtatgacgtatgacgtatgacgtatgacgtat
Db 361 ttattgttccatctgacatgacgtatgacgtatgacgtatgacgtatgacgtat
QY 421 gatgcagatagacgtcgaagaatgcatgacacccctgctgctgacgaagaatg
Db 421 gatgcagatagacgtcgaagaatgcatgacacccctgctgctgacgaagaatg
QY 481 gacctgaagactgcaataaccgctccagatgaacccagatgaatgctgagatc
Db 481 gacctgaagactgcaataaccgctccagatgaacccagatgaatgctgagatc

```

```

Db 481 GCCTTCAGACCTGCAATTAACCGTCCAGATGAACCGAGTGCAGATTAATGTGGGATC 540
QY 541 tgtttggaacagatgtgtgtgattgagttcaacgagtgctgtgtcgagaagaagtgtg 600
Db 541 TGTTCGAGAACAGATGTTGTGATGAGTCAACGAGTGTCTGTGCGAGAAAAAGTGTG 600
QY 601 ttcctgaagaatctgattccgaggaatttcctgccccagacctgtgtcttctgaaga 660
Db 601 TTCTCGAAGAAATCTGATCTCGGAGATTTTCCTGCCCCAACCTTCTGTCTTGAAGA 660
QY 661 acttcaacatctcgacattcaacggaagtgtgtacattcaagtgtgtgaatccaact 720
Db 661 ACTTCAACATCTTCGACTTTAAAGGGAAGTGTATCAAGTGTGATCAACCACT 720
QY 721 ttgattccttgacgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
Db 721 TTGATGCTTCGACCTGCGACCTGACGATGAGTTCACACAGAAAGTGAACAAGCTTGTG 780
QY 781 gaaacatctcttggagaataaagaacctagacagatgttcttactagtgtacgcgtac 840
Db 781 GAACATCTCTTGGAGATTAAGACCTTAGACAGATGATTTCTTACTAGTACGCGTAC 840
QY 841 aaaaattcgatgaagatccttaacccaactgtgtcttcaatcaatgaacagatgacc 900
Db 841 AAAAATTCGTGGAAGATCTTAACCAACCTGCTTCTTCAATCATGACAAAGAGTACC 900
QY 901 ttcactcaagaatgacgtgtatctatccctgctcaataagaatgagaatgaactgaact 960
Db 901 TTCACATCAATGATGACGTGTATCTCTGATCAAAATGAGATTAACCTGAAGACT 960
QY 961 atatatgttactacacgtgtggtggaagaatgcttggatgataatgtgtgtgtgtgtgt 1020
Db 961 ATATATTTGTATACCTGACGCTGCGCAAGAGTGTGATGATGATGATGATGATGATG 1020
QY 1021 tatcaacgagaatgctgtatctatcccaataagaatgagaatcgtgaagaacgacaa 1080
Db 1021 TATACACGAGAGTCTGTATTTACCAATACATTAATACCAAGTGTGAAGAACACCA 1080
QY 1081 aaagcatgagcagaacttcagcaatcattagaacgataaacaatgtgtctgtgaac 1140
Db 1081 AAAGCATGAGCAGAGACTTCAGCATTCATTAAGAGATTAACATGATGATGATGATGAT 1140
QY 1141 ctgcgcgtgtggaagaatgagaagaacgtgtggaagaatgagaagaatgagaagaat 1200
Db 1141 CTGCGCTGTGGAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1200
QY 1201 aggttgaagaatgagaagaatgagaagaatgagaagaatgagaagaatgagaagaat 1260
Db 1201 AGGTGAAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 1260
QY 1261 tgacctgttccagaatgctgtgaagaatgagaagaatgagaagaatgagaagaat 1320
Db 1261 TGACCTGTTCAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGAT 1320
QY 1321 tctgtaagaatgagaagaatgagaagaatgagaagaatgagaagaatgagaagaat 1380
Db 1321 TCTGTAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAG 1380
QY 1381 gtgaggttgaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
Db 1381 GTGAGGTGAAGAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
QY 1441 aaccacattgtgttacaataatataataataataataataataataataataataata 1500
Db 1441 AACCAATGTTGTAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1500
QY 1501 ttcttgacacagtcacatcattccatcaataataataataataataataataataata 1555
Db 1501 TTTTGTACACAGTCATCATTCATTCATTAACATTTGGATTAATAATAATAATAATAATA 1555

```

RESULT 2
AC003981

LOCUS	AC003981	133840 bp	DNA	PLN	18-SEP-1998
DEFINITION	Complete sequence of Arabidopsis F22013, complete sequence.				
ACCESSION	AC003981				
NID	93063438				
VERSION	AC003981.1	GI:3063438			
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.				
REFERENCE	1 (bases 1 to 133840)				
AUTHORS	Shinn, P., Buehler, E., Dewar, K., Feng, J., Kim, C., Li, Y., Sun, H., Conway, A., Conway, A., Kurtz, D., Oji, O., Shen, Y. K., Toriumi, M., Vysotskaya, V., Yu, G., Davis, R. W., Federspiel, N. A., Theologis, A. and Ecker, J. R.				
TITLE	Genomic sequence for Arabidopsis thaliana BAC F22013				
JOURNAL	Unpublished (1999)				
AUTHORS	Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-JAN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	3 (bases 1 to 133840)				
AUTHORS	Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-JAN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	4 (bases 1 to 133840)				
AUTHORS	Ecker, J. R.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-APR-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
COMMENT	On Apr 18, 1998 this sequence version replaced gi:2809231.				
FEATURES	Location/Qualifiers				
source	1..133840				
	/organism="Arabidopsis thaliana"				
	/db_xref="taxon:3702"				
	/clone="F22013"				
	/chromosome="1"				
	join(188..318,504..604,730..875)				
CDS	/note="unknown"				
	/codon_start=1				
	/evidence=not_experimental				
	/product="F22013.1"				
	/protein_id="AAC14027.1"				
	/db_xref="PID:g3063439"				
	/db_xref="GI:3063439"				
	/translation="MGRATSRRAKLADTRKELPSTLAAYRISGMEISDLTLESDLS ODITDGINSAKAAVQAAGIKOIGTIAOQOTLSMIEERANLPEISLQPVVGAAEKT SHAIGSATRKRLNMIITGKNKDE"				
	join(1384..1462,2003..2184,2418..3014,3460..3690,3802..3942,4029..4229,4332..4619)				
	/note="putative plastid RNA polymerase sigma-subunit; similar to EST gb N65838"				
	/codon_start=1				
	/evidence=not_experimental				
	/product="F22013.2"				
	/protein_id="AAC14028.1"				
	/db_xref="PID:g3063440"				
	/db_xref="GI:3063440"				
	/translation="WSSCLLPQKCPDPSSTSHFRTSCAPKHNKGSVFPOCAVST SPALITSMIDVAKLRLPESDSDSLISDRQWYTRPDGSTAEAKYLELASETLTSD DEAVVVAAGAAVALARAAYVAKADATLTKNNINILITSTADRSKRWQDETERA GILGHVAVSNGIIVSDIKITASANKESIGLSEKOEVEILLEOPSVSLAVSTROT ERKARRAKGLEKTASGIPSVKTSPPKRRKLVQAEVDNPLRLMTYSSKLLTIV EHEHLSAIDTLLEKRLQTLERSGROPTFQMAAGACVQOKSIRORIHGGTLCCKD KMTKSNRLVISTAKNTOGAKMNLQDLVDGCGRLVAGKAEKDATKGFKEFTYAHWTI KOAVKSLSDOSRMRILRPFHMEVATYRVKEARKOLYSETGHPKNEIEAIVGLSMKR				
CDS	/note="similar to receptor-like protein kinase (Ipomoea n11) (U77888)"				
	/codon_start=1				
	/evidence=not_experimental				
	/product="F22013.3"				
	/protein_id="AAC14029.1"				
	/db_xref="PID:g3063441"				
	/db_xref="GI:3063441"				
	/translation="MAVATHCTSPCHDRIRFESSDGIIGRLITRKIRNGTFLKIL PPIOSADLRRTGSSRPLSAFSGSKGIFDIPLPSKNELEKELTAPLLELVGLVA CAPIYPSADAVDALTKTCCLIKGCEIKAKIANPACAAVACIOTACNNRPDEVOA IKCGDLPENSVDNEFCVSRKRCVPRKSDIGEPAPPSVLVONENISDNGKWIY TSGLNPTFPAFDCQHEHFEHEDGNKLGNISWIKITLDSGFTSRVQKFFVDDPQPG VLYNHDNEXLYHODDWMYILSSKLENPEDEYIFVYVRGNDAMDYGAGAVYTRSSVLP NSLIPELEKAKSIGRDSFTFRTDNTCGPEVALVERIEKTVEEGERITIVKEVEIEE EKEKEVKGRTEMTLFLRLAEGFNLKQDEENFVELSKEMEFLDEIKMASVEK LFKALPIKRV"				
	/complement(7701..8633)				
	/note="syntaxin-related protein kno1le (U39451); BAC F22013 has a deletion of a cytosine at position 7887"				
	/codon_start=1				
	/evidence=not_experimental				
	/product="F22013.4"				
	/protein_id="AAC14031.1"				
	/db_xref="PID:g3063443"				
	/db_xref="GI:3063443"				
	/translation="MNDLMTKSPMSYVDLKKRAMDMAGDPDELMASTKADKMDEN LSSLEAEVYKAEEMGLISETLRLEOYHEESGVKASVSLRKISNEIVSGILRK AKGISTKLEMDKANKKIRKLTGSTPYRSTATNLGRKLEVMMEFQGLQKMESE YKTEVERITTYGCHANDMEIKITIDNAGCEETLRALIOHGKGVETVETIODR YDAKETEKSLLEHOFEDMAVMSQGDMEIDHVINASHYADANLKTAKS HQRNRKMWCIIGTIVLILIVIPILITSSFS"				
	join(10982..11218,11470..11646,11732..11896,11987..12235)				
	/note="putative thioredoxin; similar to ESTs gb T746281, gb R83933, gb N65879, emb F14466, gb N96726, gb AA042340, and emb Z18150"				
	/codon_start=1				
	/evidence=not_experimental				
	/product="F22013.5"				
	/protein_id="AAC14032.1"				
	/db_xref="PID:g3063444"				
	/db_xref="GI:3063444"				
	/translation="MTEVISTSLFLCAGCNHNRVDVDFSPVSEFGGLKRSFSLK LKSOPLRSVYFGKQIVFGSDSEFSRRAITQTLRLIGTQKMWEGKDKMWEKI USAQELVDSLTMAGDKLVVDFSPGCGGKALPKIKCOFAENNPVQFLQVYEHK SMCYSLGVHLPEFRFYGSGVSCGCTNATIKPFDALAHAGHSGSLGTGKLE EKELVALANKELNFTYTPKRPVYKAAATPNPSLPVLPENSSNDEKTLV5AGR" complement(join(13219..13280,13397..13477,13866..13932,14113..14274))				
	/note="hypothetical protein"				
	/codon_start=1				
	/evidence=not_experimental				
	/product="F22013.6"				
	/protein_id="AAC14030.1"				
	/db_xref="PID:g3063442"				
	/db_xref="GI:3063442"				
	/translation="MAKYNELAKKREKAKDRRAIHGDPILTKLITRPVPSVSGK ORKLLRKWRROKDMVEKGLVTMEDVMAQAASDESKSPKRFVSKSLKLNKLNK GKKRKNOKASGKESADCMLE"				
	16782..19871				
	/note="similar to receptor-like protein kinase (Ipomoea n11) (U77888)"				
	/codon_start=1				
	/evidence=not_experimental				
	/product="F22013.7"				
	/protein_id="AAC14033.1"				
	/db_xref="PID:g3063445"				
	/db_xref="GI:3063445"				

/translaton="MAIDRPELFYVYGCALPPEVSEFFSENEOGELILAFSDLPD
 /SNNDSMRPNATPESLVCCHTQGYHCANCVAKMLSNNEGYNLDIOQSPS
 /LOALDLSNNAESSLPLKSLNULSLKATIDVSNVSFTGTPGLGMAITGLIHVASSN
 /ESGLPELDGNATLLEVDPEGGYFEESVSSEKRNKLKIFDLSSNNGVAVYIG
 /ELSLSETLILDEYNGPELPEFEERKLRLYDIDLAVGNLTQGLPSSGLQKOLTYL
 /KONRATGLPELGLQMSLYFLDLSDMOITGLPEMVEGLKQILKONLRLDTYLP
 /SKIELEPMLEVLQMSIMSLDVLHGKNSPLKPELDVSNKRLISGIPGKLYSNTL
 /KLIFNFSSTQIEELFEESCTGLVAVTOKNHLISGIPAGSGDPLMLAKNNT
 /KIPIDALSTSLISLEIDISFNHLSLSSSTPSSNPOTGLASNNRAGTIPMDIDRP
 /SLVDLSSENFSESGTIEDRLASEFKLYSLMKNSLOLGETPLKAMHMLAVLDLSN
 /SLTGNIPDGLDCASTLEMLNENSEKLDGPIPSVSNMLPAAIDPKOLVNNLDCGVLPJC
 /SKSLALSRSKNPGRIVHNAVPELITGTSVSNMPELAGMTITPMDLSNARE
 /YIPKRPPEEMPRVLVAFORCEAGDILSHIKESNIIGMAITGLYKREAVRRLTI
 /LVAUKLMSPEPNDIDHNOEDBEEDLILREYNLGLGRRHYAYILGYVNEHEM
 /MYEAMPNGNLGALSHKDKMLRLDMLSYNVAVCYGIATLHNDKCPRIIHHDK
 /SNMLLSLLEARATDAGSLAKMLHKREIYSMAVCYGIATPEYGTTLKIDKSOIYS
 /LGVVLELVYTKMIDPDESIDIDVEMKRYKAKNKESELEVIDASIAQCKHYIEM
 /LLAIRALICLAKPIKRPISRIQVTLIAKRPKRSVOCVAGDIEFPNFSYVOLI"
 /join["22485", "22966","3045", "23140","23223", "23315","23426", "23616",
 "23701", "24112","24456", "24569","24699", "24956","25371", "25724",
 "26173", "26486","26582", "26672","26778", "26987","27131", "27258",
 "27368", "27564","27925", "28066","28322", "28474","28559", "28618",
 "28675", "28756","28806", "29002","28146", "29475","29666", "29741",
 "29856", "30082","30163", "30366","30450", "30528","30616", "30875",
 "30973", "31354")
 /note="similar to DNA dependent ATPase and helicase
 (U72937)"
 /codon_start=1
 /evidence=not experimental
 /protein_id="F22013.8"
 /product="F22013.8"
 /db_xref="PID:g3063446"
 /db_xref="GI:3063446"

[illegible]

Query Match	61.6%;	Score 958;	DB 8;	Length 133840;
Best Local Similarity	73.8%;	Pred. No. 2.1e-216;		
Matches 1541;	Conservative 0;	Mismatches 10;	Indels 537;	Gaps 4

Accession	Sequence	Position
OY 2	cagcgctccggtctggcggtggggaagattgataagcgtgaagaatgcagtgtctaca	61
Db 5342	CTCAGGATATGCTTGTTGGTGGGGAGAGATTGATAGTGTGAGCATGGCAGTAGTACACA	5401
OY 62	ttgttaacttaacctgtgcatgaccgatatgtattttctcaagtcagatgctatgg	121
Db 5402	TTGTTTCACTTCACCTTGTCATGACCGTATTCATTTTCTCAAGATGATGATGG	5461
OY 122	tagcttcggattacaagaagaagatgataatgacacttcttcgtcaagaatttaccctc	181
Db 5462	TAGGCTTGCGCATTTACAGAAAGAGGATGCATGCGCACTTCTGCTCAAAATTTTACCTCC	5521
OY 182	aatccaagtcgcatctcgaagaacactggttggaatactcgaatccttctcgaatt	241

Db	5522	AATCCAAAGTGTGATCTCAGAACAACTGGTGGGAGATCCTCAAGTCTTTATTCGCATT	5583
QY	242	caagtcagagatctcta	259
Db	5582	CAGGTGAGGATCTCTTAAGTAATCAATCACTACTAGTCTGTGTTCAATAACAATTTG	5643
QY	259	-----agggaattagatctgcatta	284
Db	5642	GTATTTGCATGATCATTGCTTTTGCATGTCAGGGGATATTATACATTGTGCCATT	5703
QY	285	ccataaagaatgagctgaaagacgctgacccgctcgtctgtgctaaactcgtgggtgt	344
Db	5702	CCATCAAGAAATGAGCTGAAAGAGCTCAACCGCTCCCGTGTGCTAAACTCGGGGTGT	5763
QY	345	ttagcttcgcgcttccttatgtcttcacgtcggagatgagagttgagtcacttaactgt	404
Db	5762	TTAGCTTGGCGTTCCTTATTGTTCATCTGCAGATGCAATTGATCACTTAATTAACCTGT	5823
QY	405	gcaitcatttgaaaggatgc	426
Db	5822	GCATCTATTGAAGGAGTCAAGTACTTCTCTGGTTGGATTCTGCATATATAGT	5883
QY	426	-----	426
Db	5882	CCCCCACCAAAACCGTTTTATTTCATTGATGTGATGCTATTAAGTGTTTGAAT	5943
QY	426	-----	426
Db	5942	GTATGCTCCATACTCTTCATATTGTGAGATGTTAAAGCTAGATGAAGTACTGTGCT	6003
QY	426	-----	426
Db	6002	GTCATGATAGAGCTGAATCAAGTATGATGATATCTTCTGATTTTTCGCCCTCAATTA	6063
QY	426	-----aggatagaactgc	439
Db	6062	TATATTCCTCCTTGTCCTACATATGAGAGCCCAATTATCTGTTAGATAGAACTGC	6123
QY	440	aaagtgcaatgccaaccctcgtctgtgagccaatgctgcgtctccttcagacctgaataa	499
Db	6122	AAAGTGCAATGCCAACCTCCTGTGAGCAATGTCGGTGCCTTCAGACCTGCAATA	6183
QY	500	ccgtccagatgaaacccgagtc	522
Db	6182	CCGTCCAGATGAAACCGAGTCCAGGTTATTAATTTTGTATATATCATTATTAAGGC	6243
QY	522	-----	522
Db	6242	AGAGTATTCGTTGCTTATGAAGAAGACCATTGATTTTGTATTATTAATTAATGCTTATA	6303
QY	522	-----cagattaaatgtgggagctcgtcttgagaaacagtgtgtatgagatccaag	575
Db	6302	TTATATAGATTAATATGTTGGGATCTGTTTGAGAACAGTGTGTTGATGAGTTCAACAG	6363
QY	576	tgtctgtgtcagaaanaaagtgtgtctctagaaaaatcgtatcctcggagaaattcctgc	635
Db	6362	TGTCTGTGTGCAAAAAAAGTGTGTCTTAGAAATCTGATCTGGAGAAATTTCTGCC	6423
QY	636	ccagaccctctgtctcttgacagaacttcaacatctcggactttaaagggaatgtgtac	695
Db	6422	CCAGACCCCTCTGTGTTCTGTACAGAACTTCAACATCTGGACTTTAAAGGGAAGTGTAC	6483
QY	656	attcaagtggtcttgaatccaaccttggatgccttcgactcgcagctgcatgagttccac	755
Db	6482	ATTTCAAAGTGGCTGGAATCCAACTTTGATGCTTCGACTGCACCTTCATGAGTTCCAC	6543
QY	756	acagaagtgtaacaacaagctgtgtgnaaacatctcttgagaaataaagacccctgaagt	815
Db	6542	ACGAAGGTGACAAACAAGCTGTGTGGAACAATCTTTGGAGATTAAGAACCTTAGACAGT	6603
QY	816	ggattcttactagctcagccgcbacaaaattcgtgcaagatcctaaccaaccggtgt	875
Db	6602	GGATTTCTTTACTGTAGCAGCCGTCACAAAATATGTCGAAGATCCTTAACCAACTGGTGT	6663

QY	876	ctctacacatattgcaacagagctaccttccactacatacaaaatgac-----	919
Db	6662	CTCTACATCATGTGCAACAGAGTACCTTCACCTATCAAGATGACTGTTAATACATCTTGTGA	6721
QY	919	-----	919
Db	6722	TCCACCACAGAGGAGAGTCTGATAGTGTGCTATCTTCTTGCGCTAACAGCTTTC	6781
QY	919	-----	919
Db	6782	ATATGATTTTTCAGGATATTCCTGTATCAAGATATGAAGATTAACCTTAACACTATAT	6841
QY	965	atttgctactacagctgggcaaacgcatgctcttgagatcgatgctgctgctgata	1024
Db	6842	ATTGTGTACTACCGTGGGCGCAACGATGCTTGGGATGATATGCTGTGCTGATGATA	6901
QY	1025	cacgagaaagctctgtatcccaataagcatataccagaactcgaanaagcagcaaaag	1084
Db	6902	CACGGAAGTCTGTATTACCATAGCATATTACCGAAGCTCGAANAACAGCAAAAG	6961
QY	1085	cataggcaagagacttcagcacatcattagaacggaataacacatgctgcttgaacctgc	1144
Db	6962	CATAGGCGAGACCTTCAGCACATTCATTAGAACGGATTAACACATGTGTGCTTGAACTGC	7021
QY	1145	gctcgtgagagaatttgaagaacagctggaagaagctggaagataactgtaaaagagt	1204
Db	7022	GCTCGTGGAGAGCAATTGAGAGACAGCTGGAAGAGCTGGAAGATATCTGTAAGAGGT	7081
QY	1205	tgaagaataaagaagaagagctagaagaagctggaagaagctcgctgagactgagatgac	1264
Db	7082	TGAAGAGATTAAGAGAGAGGTAGAGAAAGAGAGTGGAGAGGTCCGTGTGAGCTGAGATGAC	7141
QY	1265	cttgctccaagagattgctgtagaagatttaagaaactgaactgaagaagaagagaaatttcgt	1324
Db	7142	CTTGTTCCAGAGATTGGCTGGAAGATTTAATGAACCTGAAGCAAGCGGGAGATTTTCGT	7201
QY	1325	gagaaagctaaagaagaagaatgagatcttcttgatbagaatacaaatggaagcaagtga	1384
Db	7202	GAGAGAGTTAAGTAAGAAGAGATGCAATTTTGGATAGATCAAAATGSAAGCAAGTGA	7261
QY	1385	gattgaaaaaatgtttgggaaagcttgcacaatcagaaggtcaggtgagaacaagaacc	1444
Db	7262	GGTTAAAAAATGTTGGGAAAGCTTTCACATCGAAGAGGTCAAGGTGAAGAACAGAAACC	7321
QY	1445	accatttgttgaacaacatattatcacatgctgttgcgttcataataagtaatttt	1504
Db	7322	ACCATTGTTGTAACAACATATATTATACATACTGTGTTGCTTCATATTAAGTAAATATTTT	7381
QY	1505	tgtacaagctcatcatcattccataaacatitgataaaaaa	1552
Db	7382	TGTACACAGTCAATCATCATTCATTAACAATGGATACAGAAACACAA	7429
RESULT	3		
LOCUS	ATU39452/c	2884 bp	DNA
DEFINITION	Arabidopsis thaliana		PLN
ACCESSION	U39452		10-FEB-1996
VERSION	91184166		complete cds.
KEYWORDS	GI:1184166		
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

	MEDLINE	96140640
	REFERENCE	2 (bases 1 to 2884)
	AUTHORS	Lukowitz W.
	TITLE	Direct Submission
	JOURNAL	Submitted (26-Oct-1995) Wolfgang Lukowitz, Lehrstuhl fuer Entwicklungs-genetik, Universitaet Tuebingen, Spemannstrasse 37-39, D-72076 Tuebingen, Germany
	FEATURES	location/Qualifiers
	source	1..2884
		/organism="Arabidopsis thaliana"
		/strain="Columbia"
		/db_xref="taxon:3702"
	gene	320..1452
		/gene="knolle"
	CDS	520..1452
		/gene="knolle"
		/note="syntaxin-related"
		/codon_start=1
		/protein_id="AAC49163.1"
		/db_xref="PID:g1184167"
		/db_xref="GI:1184167"
		/translation="MNIDMTKSPMSYVDLTKKAAMDKEAGPDFLEMASTRADKNDEH LSPEEFAEVYAKWGLISETLRIEQYHESGVHKAEVSKLRKINSEIVSLRISVSLRA ASISKSLKEMDKAKETRLGSRGPYRSRYAINGIKRKIKVMVEFQGLKKXSS KEIYERIRFIYTGHEHANDEITEKITDDNKGEPFLTRAQEGRRKVKVFETVLODD CYAAEKIEKSLHEQFLDMAAVVESQGDMDIEIHVIYNASHIVADANELKTAKK HORNSRWKCIOIIVLLILIIIVIPPIITFSFS"
	BASE COUNT	833 a 569 c 582 g 900 t
	ORIGIN	
Query Match	59.3%; Score 922; DB 8; Length 2884;	
Best Local Similarity	91.4%; Pred. No. 5.8e-208;	
Matches 1026; Conservative	0; Mismatches 5; Indels 91; Gaps 1	
Qy	522	cagatataatgttgaggatctgcttttggagaacagtgttgttaagattcaacagatgtcgt 581
Db	2845	CAGATTAAATGTGGGACATCTGTTTGGAAACAAGTCGTTTGAAGACTTCAACAGACTGTGCT 2786
Qy	582	gtgtcgagaaaaaagatgtgttccttagaaaaatcgtactcgtcgaaattccctgcccgac 641
Db	2785	GHTCGACCAAAGAAGTGTTGTTCTTAGAAAAATCGATCTCGAGAATTTCTCGCCCGAC 642
Qy	642	ccttcgtgtccttgacaagaactccaacatcctggaccttaaagggaagtgttacattaca 701
Db	2725	CCTTGTGTTCTTGNACGAACCTTCAACATCTGGACTTTAAAGGGAATGGTACATTACA 2666
Qy	702	aatgtcgtgaatccaacctltgatgccttcgcagctgcagctgcagatltcacacagaa 761
Db	2665	AATGCGCTTGATCCAACACTTTGATGCTTGATGCCACTGCATGAGTGCACACAGAA 2606
Qy	762	ggtgacaacaagcttgtttggaacaacatcctcttggagaataaagacccttagacaagyattc 821
Db	2605	GGTGACAACAAGCTTGTTGGAAACATCTCTGGAGAAATAAACCCTTAGACAGTGGATTG 2546
Qy	822	ttaactagatgcggccgtacaaaaattgtgtcaaagatccctaaccaacctggtgtctctac 881
Db	2545	TTTTCTAGGTCAGCCGTACAAAATTGTGTGTAAGATCCTTACCAACCTGTGTCTCTAC 2486
Qy	882	aatcatgacaacagattaccttcaatcaaatcaaatgact----- 919
Db	2485	AATCATGACAACAGATACCTTCACTATCAAGATGACTGTAATAACATCTGTATCACAC 2426
Qy	919	----- 919
Db	2425	CACGAGGAGAGACTTGTAGTACTGCTATTCTTTTGCCCTAACACACTTTCATATGA 2366
Qy	919	-----ggtatccgtgcatcataaagatagaaataaacctgaaagactatatattgt 970
Db	2365	TTTTTTCAGAGRTATCCCTGTCATCAAGATAGAGAAATAAACCTGAAGACTATATATTGT 2306
Qy	971	atactaccgttgaggcaaacgactgttggaatgatatgtgtgtgtgcaattggtataaacag 1030

Db 2305 AACTACCGTGGGCGAAGATGCTTGGATGATATGCTGCTGATGATACAGAG 2246
 QY 1031 aagcttctgattacccaatagatataccgaatctgcaaaaaacgcaaaagcattagg 1090
 Db 2245 AAGTTCTGATTAACCAATAGCATTATACGAACTCCAAAAACACCAAAACCATAGG 2186
 QY 1091 cagagactcagacattcattatagacagatacacatgctgctgaactgctgctcgt 1150
 Db 2185 CAGAGACTCAGACACATTTCATTAGAACGATTAACCAATGCTGCTGCTGCTGCTGCT 2126
 QY 1151 ggaaggaattggaaggaagcaagtggaaggaagtgaaggaatacgttaaaaggttgaaga 1210
 Db 2125 GGAGGAATGGAAGACAGTGGAAAGAGTGAAGAGTAAATCGTAAAGAGGTTGAGA 2066
 QY 1211 gataaagaaggaaggaaggaaggaagtggaaggaagtggaaggaagtggaaggaagtgga 1270
 Db 2065 GATTAAGAAGAGGATGAGAAAGAGTGGAAAGAGTGGAAAGAGTGGAAAGAGTGGAAAG 2006
 QY 1271 ccagagattgctggaaggaatttaactgaactgaacagaggaaggaattcgtggaaga 1330
 Db 2005 CCAGAGATTGGCTGAGAGGATTTAATGAACTGAGCAAGACAGAGGAGATTTCTGAGAGA 1946
 QY 1331 gtaagtaaaaggaaggaaggaaggttttggatgagatcaaaatggaaggaagtgagggtga 1390
 Db 1945 GTTAAGTAAAGAGAGATGAGGATTTTGGATGAGATCAAAATGGAACAGAGGAGGTGA 1886
 QY 1391 aaatctgttgggaaggttgcacatcaggaaggtcaggtgagaagaagaacaccatt 1450
 Db 1885 AAAATGTTGGGAAGAGTGGCAATCAGGAAGAGTGAAGTGAAGCAACCAACCACTT 1826
 QY 1451 gttgacaactataatatacactctgtctgctggtcattaaagtaataattttgttaca 1510
 Db 1825 GTTGACAACTAATATATATACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1766
 QY 1511 cagtcacatcattccatacaactggaataaaaaa 1552
 Db 1765 CAGTCATCATTCATTCATTAACATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1724

RESULT 4
 NTU34817 1589 bp mRNA PLN 15-JUN-1998
 LOCUS NTU34817 1589 bp mRNA PLN 15-JUN-1998
 DEFINITION Nicotiana tabacum violaxanthin de-epoxidase precursor (TVDEL),
 mRNA, complete cds.
 ACCESSION U34817
 NID U34817.1 GI:1463122
 VERSION U34817.1
 KEYWORDS common tobacco.
 SOURCE Nicotiana tabacum
 ORGANISM

REFERENCE
 AUTHORS Bugeo, R.C., Hieber, A.D. and Yamamoto, H.Y.
 TITLE Xanthophyll cycle enzymes are members of the lipocalin family, the
 first identified from plants
 J. Biol. Chem. 273 (25), 15321-15324 (1998)
 MEDLINE 98288256
 REFERENCE 2 (bases 1 to 1589)
 AUTHORS Bugeo, R.C. and Yamamoto, H.Y.
 TITLE Direct Submission
 JOURNAL Submitted (25-AUG-1995) Robert C. Bugeo, Plant Molecular
 Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI
 96822, USA

FEATURES
 source
 Location/Qualifiers
 1..1589
 /organism="Nicotiana tabacum"
 /strain="Xanthi"
 /db_xref="taxon:4097"
 /feature_type="leaf"
 1..1589
 gene

CDS
 /gene="TVDEL"
 42..1478
 /gene="TVDEL"
 /codon_start=1
 /product="violaxanthin de-epoxidase precursor"
 /protein_id="AAC50031.1"
 /db_xref="PID:g1463123"
 /db_xref="GI:1463123"
 /translation="MALAPHSNFIANHETIKYVGSKLPGHKFSGMGMEDYFSGIVV
 KISSRIRIPRYFRKSPRIOCGLDSRGLQFSHGKHNLSPAHSINONYPKNGSGCKRFX
 DVANLWEMKGOFAKTAIVAFILISVASKADAPALAKCTLLAECHLAKCISNPA
 CANAVACLOTQCNRRDRETCQKCCDLENSVDFEENCAVSRKCVPRKSDVDPEV
 PDPSTVVOFDMKDFSGKWEITRGSLNPFDFADQLEHFEHTEENKLVGNLSMRTPD
 GGFTRSAQKTVQDPKYPGLIYNHNDNYLLYQDDWYILSSKNSPEDYIFVYKGR
 NPAMDYGGSVLYTRSAVLPESILPELOTAOKVGRDPEFTIKTDNCGEPPELYERL
 EKVEGERTLIKVEEIEEVEYKVRDEKVTFLFKLEFGEFRELQDRENLRLSKKE
 MDVLDLKKMEATEVEKFLGRALPIRKLR"

transit_peptide 42..443
 /gene="TVDEL"
 44..1475
 /gene="TVDEL"
 /product="violaxanthin de-epoxidase"

BASE COUNT 500 a 262 c 366 g 461 t
 ORIGIN

Query Match 40.1%; Score 624.2; DB 8; Length 1589;
 Best Local Similarity 74.3%; Pred. No. 1.2e-137;
 Matches 820; Conservative 0; Mismatches 268; Indels 15; Gaps 2;

QY 364 ttgttcacatcgtgagatgagtggaattgagcactaaacttggcatgcttatgaaggat 423
 Db 424 TTGCTTCAAAAGCTGATGCGGTTGATGCTCAAACTTGATCTGCTTACTGAAAGAGT 483
 QY 424 gcaagataaactcgcgaagatgcatgccaaccctgctgcaagcaccattgctgagc 483
 Db 484 GCAGTTAGAGCTTGGAGAGTGCAATTTGCAACCCGCAATGCGACACTAATTTCCCTGTC 543
 QY 484 ttgaaccctgcagtaacgctccagatgaaccgagtgccagatnaatltggtggatctgt 543
 Db 544 TCAGACTTGCACCAATTAACCTGACGAAACGAAATGCAATTAATATGTTGATTTGT 603
 QY 544 ttgaagacagtggtgtgtgagtgatccaagagtggtgctgtgtgcaagaaaaagtgcttc 603
 Db 604 TTGAAAACAGTgctgtagcagagttcaatgaatgtgctgctcccgaaagaaatgtgtac 663
 QY 604 ctgaagaaatcgtatctcggaagaatttctgcccagacccttctgttctgtacagact 663
 Db 664 CTCGTAATCTGATGTGTGACTTCTCTGACCTGATCCGATCCAGTGGTCTGTCCAGAGT 723
 QY 664 tcaacatctcgagacttgaaggaagtggtacattacaagtggttgaatccaacttg 723
 Db 724 TTGACATGAAAATTTTACCGGGAATGTTCAATTAATCGCGGTTTAAATCCACTTTTG 783
 QY 724 atgctctgcacgcagctgcatgagttccacacagaagtggaacaagaactgttgaa 783
 Db 784 ATGCTTTGATTTGCCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 QY 784 acatctcttggaagaaagacctaagacagtgagattcttactaagtcagccgtacaaa 843
 Db 841 ATTATCTTGGGAGTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 QY 844 aatctgtgaagatcctacaacacactggttctctcaatcacaatcagaagaatcttct 903
 Db 901 AATGCTGCAAAATGCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 QY 904 actacaagaagatcgtgataatcctgcatcaagaagatagaagaataaactgaagactata 963
 Db 961 TCTACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 QY 964 tattgttactacccgtgtggcgaagatgcttggatgataatgtgtgtgagttgatat 1023
 Db 1021 TATTTGTACTATAAGGGGAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1080

QY 1024 acaagagaagttcgtatccatagcattaccagaactcgaagaagcaaaaa 1083
 DB 1081 ACACAGAAGAGCGATTTCCTGAAAGCATATACCGGAGTTCACACCGCAGCTCAAA 1140
 QY 1084 gcatagagagagactcagcactcattagaacgagatacacatggtcctgaacctg 1143
 DB 1141 AAGTTGGCGTGATTTTCAACCATTCATTAATAAGACATACATGTGGCCCTGCAACCTC 1200
 QY 1144 cgtcgtgagagaagattgagaagacagtggagaagaggltgaagaatacgtgaagaag 1203
 DB 1201 CCTTGTGTGAGAGGTGGAGAGAAGAGCAAGAGAGAGAGCATATTAAGAGAG 1260
 QY 1204 ttgaagaatgagaagaaggtgaagaagagagtgaggaagagtgctgtagacgcgagatga 1263
 DB 1261 TTGAGAGATGAGAAAGAAAGATGAGAGAGGTGAGAGATTAAGAGAG-----TCA 1308
 QY 1264 ccttgccagagattggtcgaagatttaagactgaacgaacgaagcagagagattcg 1323
 DB 1309 CCTTATTCAGTAAGTCTTTGAAGTTTAAGAGCTCCAAAGAGATGAAGAGACTTCT 1368
 QY 1324 tgaagagattgaagaagaagatgaggttttggatgagatcaaaatggaagcaagt 1383
 DB 1369 TAAGAGAGCTGAGCAAAAGAAATGATGTTTGGATGAGACTTAAGATGAAGCAACTG 1428
 QY 1384 aggttgaaaatgtgttggaagagcttgccaatcaggaaggtcaggttagaagaagaac 1443
 DB 1429 AGGTGAAAGAACTTTTGGCGCTGCTTTACCAATAGAGAAATTAAGTAACTATTTTAA 1488
 QY 1444 caccattgtgtacaactatat 1466
 DB 1489 AACTATCAACATATATACTACAT 1511
 RESULT 5
 LSU31462 1760 bp mRNA PLN 23-JUL-1996
 LOCUS Lactu saliva violaxanthin de-epoxidase (VDE1) precursor, mRNA,
 DEFINITION complete cds.
 ACCESSION U31462
 NID 91438874
 VERSION U31462.1 GI:1438874
 KEYWORDS
 SOURCE Lactu saliva.
 ORGANISM Lactu saliva.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Asterales; euasterids II; Asteriales; Asterales; Lactu saliva.
 REFERENCE 1 (bases 1 to 1760)
 AUTHORS Bugos, R.C. and Yamamoto, H. Y.
 TITLE Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6320-6325 (1996)
 MEDLINE 96270536
 REFERENCE 2 (bases 1 to 1760)
 AUTHORS Bugos, R.C.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUL-1995) Plant Molecular Physiology, University of
 Hawaii, 1390 Maile Way, Honolulu, HI 96822, USA
 FEATURES
 source
 /organism="Lactu saliva"
 /strain="Romaine"
 /db_xref="taxon:4236"
 transit_peptide 235..609
 /gene="VDE1"
 /gene="VDE1"
 /gene="VDE1"
 /codon_start=1
 /product="violaxanthin de-epoxidase precursor"
 /protein_id="AAC49373.1"
 /db_xref="PID:91438875"
 /db_xref="GI:1438875"

/translation="MALSLHTFLCKEALNLVARSPONENFRHSGOPTNIMKIR
 SNNGYENSRFLSYKTSFSSDSSHKRKSQICSIDSFIEQIFDKRMFTLLEKO
 RHOPIALIVATCEVIVPRVDADALKTCAKLEKRIELAKIAPSCAANVACIO
 TGNRRDEDECOIKCGDLFENSVDPOFEKASRKKCPKRSVDGEPPVDRNAVON
 ENKDKSGKWTITSGNLPTFDFDCOLIEFHEMDKLVGNITWIKTLIDGFEFRSAV
 QFEVDPPDLGALYNHNEFLHYDDWVILISQLENNPDDITFYIYGRNDAMDYGK
 SVITRSPILPESIIIPNLOKAKSVGRDFNFIITDNGCGEPPLVERLEKTAEGEK
 LIKEAVEIEEVEKEVEREDFEMTLFLRLELFGKELQODENFVRELSEKEEILN
 ELQMENTEVEKLFGRALPIRKLR"
 235..1656
 /gene="VDE1"
 /gene="VDE1"
 /gene="VDE1"
 /product="violaxanthin de-epoxidase"
 mat_peptide
 /gene="VDE1"
 /product="violaxanthin de-epoxidase"
 BASE COUNT 557 a 319 c 387 g 497 t
 ORIGIN
 Query Match 39.3%; Score 611.8; DB 8; Length 1760;
 Best Local Similarity 72.5%; Pred. No. 9, 9e-135;
 Matches 806; Conservative 0; Mismatches 302; Indels 3; Gaps 1;
 QY 324 ttgctaaacctcgtgggtgttttagcttgccgttccattatgttccactcagatga 383
 DB 550 TTCATACAAATGGGTATCGATTGGTTGACAAATTTGTTATCTTCCAGAGTGAATGCC 609
 QY 384 gttagtgactiaaactgtgcatgtctttagaagagatcagagatgaactcgaag 443
 DB 610 GTTAGTGCTCTTAACACTTGTGCTTGTACTCAAAATGCAAGATGAGCTTGCAGAA 669
 QY 444 tgcattgccaacctgcctgtgcaagcaatgctgcgtgcttcaagacctgcaaatcgt 503
 DB 670 TGTATAGCAAAACCATCTTGTGGCGCAAAAGCTTGCCTGTGTACAGACTTGCAACATCGT 729
 QY 504 ccagatgaacccaggtgcagataaaltgtyggagatctttagaagaagatgtgtgtat 563
 DB 730 CCGTACGAGACGSAATGTCAGATTAATGTGTAATCTTGTTCGAAACAGTGTGTGGAC 789
 QY 564 gattccaacgaatgtgctgtgctgcaagaagaatgtgttcttcaagaatcgtatcgtga 623
 DB 790 CAATTCACGAGATGTGCGGCTTCCCGAAAGAAATGTGTGCCCGGAATGAGATGGGG 849
 QY 624 gaattccctgccagacccttgcctgttctgttacaagaactcaacatcgtgacttaac 683
 DB 850 GAATTCGCCGTCCCGATCGTAATGAGTGTTCATAATTTTACATGAAGAACTTTAGT 909
 QY 684 ggaagtggtatatacagaagtgcttgatccaacttgatgcttgcgtcagctcgaagtg 743
 DB 910 GGAAGTGTATATACAAAGTGTTAATCTCATTTGATGATTTGATTTGCAACTT 969
 QY 744 catgaattccatccagaaggtgacacagaactgttgaagaactcttcttgagaataaag 803
 DB 970 CATGAGTTTCATATGGAAGAATG---ATAAATGTGTGGAACCTTAACATGGCGCATAAA 1026
 QY 804 accctagaagatgagatcttactaggtccagcgtacacaaatcgtgcaagaatccctaac 863
 DB 1027 ACTTTGGATGGGTCTTCTTACTGATCTCTGTGCAAAATTTGTTCAGAGATCCGAT 1086
 QY 864 caactcgtggtctctacacatgacagaagatccctcactcaagaatgagatgtat 923
 DB 1087 CTTCCTGGAGCACTTTATATATATGACATAGATTTCTTCACTACCAAGATGATGTTAC 1146
 QY 924 atcctgtcaataagataagaataaactgaagaactatattgtatactacgttgg 983
 DB 1147 ATATATCTTCCAAATCGAAACCAACCCGATGATTCATATTTGTAATACCGAGGT 1206
 QY 984 cgaagaagatcgtgagatgagatgtgtgtgagttgagttgacagagaagatcgtgtat 1043
 DB 1207 CGAAACGACGATGGATGATGATCGTGGTCCGATCTACCCGAGAGCCGACACTC 1266
 QY 1044 cccaatgagatcctacagaactcgaagaagaagaagaacataagcagaagatcagc 1103
 DB 1267 CCGGATTCGATATCCCAACCTACAAAGAGCAAAATCGTGGGTGCGAGACTTTAAC 1326

[illegible]

Query Match	Best Local Similarity	Score	DB	Length
Matches 164; Conservative	50.0%	Pred. No. 0.0032;	Mismatches 162; Indels	Gaps 1
QY 1080	aaaagcatalgcsagagacttcacgscatcattgaacgagataacacatggtgtccgaa	1139		
Db 894	AGAAAGAGAAAGTGTGAAGGGGACAGAAAGAGAAAGTGAAGGGGACAGAGTGAAG--AAATA	951		
QY 1140	ccctggtctgtgagagaaattgagaagacagtcggaagagagtgaaagatatcgtataa	1199		
Db 952	GAAGGACACAGAAAGAGAAAGTGAAGGGGACAGAAAGAGAAAGTGAAGGGGACAGAGTGAAG--AAATA	1011		
QY 1200	gaagttgaagaataaagaagaagtgagaagaagtgagaagagtcgtgaagactgag	1259		
Db 1012	GAAGTGAAGAGGACAGAGAGAGTGAAGGGGACAGAAAGTGAAGGGGACAGAA	1071		
QY 1260	atgacctgtctcagagatgtgctgtaagattaatgacacggaacgaagagagagaaat	1319		
Db 1072	GAGGAAGTGAAGAGGACAGAGAGTGAAGGGGACAGAAAGAGAGTGAAGGGGACAA	1131		
QY 1320	ttcgtgagagagttaagtaagaagagatcgagattttgattgagatcaaaattgaaaga	1379		
Db 1132	GAAGGAGAAATAGCAAGGAGACAGAGAGAGTGAAGGGGACAGAGAGTGAAGAGG	1191		
QY 1380	agtgaagttgaaaaattgttttggaag	1407		
Db 1192	ACAGAGAGTGAAGTGAAGGGGACAG	1219		

RESULT 9

AP000067/c

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

100000 bp

DNA

PRT

17-FEB-1999

Homo sapiens genomic DNA of Bp11.2 senescence gene, segment 3/19.

Complete sequence.

AP000067

94579988

AP000067.1

GI:4579988

HtG.

Homo sapiens DNA.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (sites)

Isumura, M., Ikegawa, S., Kinjo, T. and Nakamura, Y.

TITLE DNA sequence analysis of a 1.9-Mb region on chromosome 8p11.2
 JOURNAL Published Only in Database (1999) in press
 REFERENCE 2 (bases: 1 to 100000)
 AUTHORS Hirakawa, M., Yamaguchi, H., Imai, K. and Shinada, J.
 TITLE Direct Submission
 JOURNAL Submitted (12-FEB-1999) to the DDBJ/EMBL/GenBank databases. Mika Hirakawa, Japan Science and Technology Corporation (JST), Advanced Database Department, 5-3, Yonban-cho, Chiyoda-ku, Tokyo 102-0028, Japan (E-mail: mika@tokyo.jst.go.jp, Tel: 81-3-5214-8491, Fax: 81-3-5214-8470)

COMMENT

This sequence is conducted by Japanese Foundation for Cancer Research as a JST sequencing team.
 Principal Investigator: Yusuke Nakamura Ph.D
 Phone: +81-3-5449-5372, Fax: +81-3-5449-5433,
 yusuke@jst.go.jp
 The sequence is submitted by Human Genome Sequencing in ALIS project of JST
 Japan Science and Technology Corporation (JST)
 5-3, Yonban-cho, Chiyoda-ku, Tokyo, 102-0028 Japan
 For further information about this sequences, please visit our sequence archive Web site (<http://www.alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www.alis.tokyo.jst.go.jp.

FEATURES

Source

1..100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8p11.2"
 66183..66318
 /standard_name="D8S571"
 /note="1264/895: The location is between each flanking site of PCR primers."
 /db_xref="GDB:214930"

STS

BASE COUNT 24961 a 22749 c 24235 g 28036 t 19 others
 ORIGIN

Query Match 3.5%; Score 54.2; DB 9; Length 100000;
 Best Local Similarity 48.8%; Pred. No. 0.005;

Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1143 gcgcctcgtgaggaattggaagacagctggaagaagttaaaagataatcgtaaaagag 1202
 Db 89163 GAGATTGAGAAAAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 89104
 QY 1203 gttaagaagataagaagaagatggaagaagtgcgtgaagactgagatg 1262
 Db 89103 GAATTAAGATAGAGAGAGAGAGAGAGATGAGAGAGAGAGAGAGAGATGAGAG 89044
 QY 1263 accctgtccagagattgctgaaggaatttaatgaactgaagaagaagaatttc 1322
 Db 89043 GAGACAGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAG 88984
 QY 1323 gtgagaagataagaagaagatggaagattttggaatgagatacaaaatgaagaagt 1382
 Db 88983 ATGGAAGAAAAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAG 88924
 QY 1383 gaagttgaaaatattgtttgggaagctttgccaatcagaagaagtcaggtgaacaacga 1441
 Db 88923 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGAGAGAG 88865

RESULT 10
 AC004670/c

LOCUS AC004670 200000 bp DNA HTG 19-JAN-1999
 DEFINITION Homo sapiens chromosome 4, WORKING DRAFT SEQUENCE, 21 unordered
 pieces.

AC004670

AC004670

AC004670

AC004670

AC004670

AC004670

AC004670

AC004670

AC004670

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases: 1 to 200000)
 AUTHORS Stone, N.E., Schmutz, J.J., Cox, D.R. and Myers, R.M.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases: 1 to 200000)
 AUTHORS Stone, N.E., Schmutz, J.J., Cox, D.R. and Myers, R.M.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAY-1998) Department of Genetics, Stanford Human Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
 On Jan 19, 1999 this sequence version replaced g1:3128159.
 NOTE: This is a "working draft" sequence. It currently consists of 21 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1705: contig of 1705 bp in length
 * 1706 3345: gap of unknown length
 * 3346 4915: contig of 1570 bp in length
 * 4916 6555: gap of unknown length
 * 6556 8554: contig of 1999 bp in length
 * 8555 10194: gap of unknown length
 * 10195 11813: contig of 1619 bp in length
 * 11814 13453: gap of unknown length
 * 13454 15141: contig of 1688 bp in length
 * 15142 16781: gap of unknown length
 * 16782 19060: contig of 2278 bp in length
 * 19060 20698: gap of unknown length
 * 20699 23780: contig of 3082 bp in length
 * 23781 25420: gap of unknown length
 * 25420 27561: contig of 2141 bp in length
 * 27561 29199: gap of unknown length
 * 29199 31694: contig of 2494 bp in length
 * 31694 33332: gap of unknown length
 * 33333 35756: contig of 2424 bp in length
 * 35757 37395: gap of unknown length
 * 37396 41867: contig of 4472 bp in length
 * 41868 43506: gap of unknown length
 * 43507 46908: contig of 3402 bp in length
 * 46909 48547: gap of unknown length
 * 48548 55405: contig of 6858 bp in length
 * 55406 64693: gap of unknown length
 * 64694 66332: gap of unknown length
 * 66333 75103: contig of 8771 bp in length
 * 75104 76742: gap of unknown length
 * 76743 85872: contig of 9130 bp in length
 * 85873 87511: gap of unknown length
 * 87512 102749: contig of 15238 bp in length
 * 102750 104388: gap of unknown length
 * 104389 117449: contig of 13061 bp in length
 * 117450 119088: gap of unknown length
 * 119089 143702: contig of 24614 bp in length
 * 143703 145341: gap of unknown length
 * 145342 170412: contig of 25071 bp in length
 * 170413 172051: gap of unknown length
 * 172052 200000: contig of 27949 bp in length.

location/Qualifiers
 1..200000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"

BASE COUNT 41331 a, 41412 c 41700 g 42768 t 32789 others
 ORIGIN

Query Match 3.4%; Score 53.2; DB 34; Length 200000;
 Best Local Similarity 50.4%; Pred. No. 0.0088;
 Matches 130; Conservative 0; Mismatches 128; Indels 0; Gaps 0;


```

repeat_region      13109..13328
                    /rpt_family="Charliel"
repeat_region      complement(13346..13492)
                    /rpt_family="AluJo/FRAM"
repeat_region      14237..14896
                    /rpt_family="LIMB2"
repeat_region      14928..15068
                    /rpt_family="FRAM"
repeat_region      complement(16993..17105)
                    /rpt_family="MLTIC"
repeat_region      17181..17594
                    /rpt_family="MLTIC"
repeat_region      17914..18278
                    /rpt_family="LIMB4"
repeat_region      complement(19315..19336)
                    /rpt_family="(CA)n"
repeat_region      20388..20447
                    /rpt_family="MIR"
repeat_region      20898..21204
                    /rpt_family="AluYa5"
repeat_region      complement(21454..21752)
                    /rpt_family="MLT2FB"
repeat_region      complement(21753..21949)
                    /rpt_family="AluJo"
repeat_region      21950..22095
                    /rpt_family="AluY"
repeat_region      22096..22380
                    /rpt_family="AluY"
repeat_region      22381..22404
                    /rpt_family="AluY"
repeat_region      22405..22563
                    /rpt_family="(TAA)n"
repeat_region      complement(22564..22695)
                    /rpt_family="AluJo"
repeat_region      complement(22696..22753)
                    /rpt_family="MLT2FB"
repeat_region      complement(22754..22917)
                    /rpt_family="AluYa8"
repeat_region      complement(22918..23222)
                    /rpt_family="AluY"
repeat_region      complement(23223..23369)
                    /rpt_family="AluYa8"
repeat_region      complement(23370..23482)
                    /rpt_family="MLT2FB"
repeat_region      complement(24361..24670)
                    /rpt_family="AluX"
repeat_region      25611..25753
                    /rpt_family="MER91A"
repeat_region      26058..26533
                    /rpt_family="MER91A"
repeat_region      26535..26840
                    /rpt_family="AluSq"
repeat_region      27581..27617
                    /rpt_family="MLT2CB"
repeat_region      complement(28183..29199)
                    /rpt_family="LIMse"
repeat_region      30034..30098
                    /rpt_family="L2"
repeat_region      30241..30521
                    /rpt_family="AluSc"
repeat_region      30522..30559
                    /rpt_family="(CAA)n"
repeat_region      31138..31165
                    /rpt_family="Al_rich"
repeat_region      31421..31780
                    /rpt_family="MLT1A1"
repeat_region      complement(31817..31885)
                    /rpt_family="(CA)n"
repeat_region      complement(33652..33918)
                    /rpt_family="AluX"
repeat_region      37861..37925
                    /rpt_family="GC_rich"
repeat_region      38035..38056

```

```

repeat_region      /rpt_family="GC_rich"
repeat_region      38108..38132
                    /rpt_family="GC_rich"
repeat_region      38296..38331
                    /rpt_family="GC_rich"

Query Match      3.4%; Score 52.8; DB 11; Length 169931;
Best Local Similarity 53.7%; Pred. No. 0.011;
Matches 132; Conservative 0; Mismatches 112; Indels 2; Gaps 1;

QY 1149 gtgagagaaatcgaagacagtcgaagaaggtgaagaataatcgtaaagaggttga 1208
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2401 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1209 gagataagaagaagtagaagaagaagtggaagagtggtcgtgaagacgtgaagcttg 1268
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2459 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1269 ttccagagatcgtgaagatttaactgaactgaagcaagacgaagagatttcgtga 1328
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2519 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1329 gagttaagtaagaagaagatgagtttttgatgatacaaaatgaagcaagtgaagtt 1388
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2579 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1389 gaaaaa 1394
    |||||
Db 2639 GAAGAA 2644

```

```

RESULT 12
HSAU9612      153053 bp      DNA      HTG      11-NOV-1998
LOCUS      Homo sapiens chromosome 17 clone PAC 1577 map p11, WORKING DRAFT
DEFINITION      SEQUENCE, in unordered pieces.
ACCESSION      AJ009612
NID      93581972
VERSION      AJ009612.1 GI:3581972
KEYWORDS      HTG; HTGS; PHASE1.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE      1 (bases 1 to 153053)
AUTHORS      Radclott,U., Hennig,S., Ramser,J., Francis,F., Steffens,C.,
            Klein,M., Seranski,P., Poustka,A., Reinhard,R. and Lehrach,H.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-1998) Steffen Hennig, MPIMG, Abt. Lehrach, Max
            Planck Institut fuer Molekulare Genetik, Innesstrasse 73, Berlin,
            14195, Germany
COMMENT      HTGS-Phase1: Unordered fragments separated by 800 N nucleotides.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

```

```

FEATURES
    source      location/Qualifiers
                1..153053

```

```

BASE COUNT      38825 a 34160 c 33964 g 39097 t 7007 others
ORIGIN
    1..153053
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="17"
        /clone="PAC 1577"
        /map="p11"
        /note="region between markers D17S71-D17S58"

```

```

Query Match      3.4%; Score 52.8; DB 34; Length 153053;
Best Local Similarity 53.7%; Pred. No. 0.011;
Matches 132; Conservative 0; Mismatches 112; Indels 2; Gaps 1;
QY 1149 gtgagagaaatcgaagacagtcgaagaaggtgaagaataatcgtaaagaggttga 1208

```

[illegible][illegible]

BASE COUNT	923 a	711 c	945 g	582 t
ORIGIN	QGFDLMNKEFLYYGRRINLNDGLKHEHREITDPLIDPDSMEPTWAKSEQQRVKRGTS PRPPEGIGTSQLDDDDLKETGFIHLTTTNGSAAAGPGRSLAF"			

Query Match	3.3%;	Score 50.6;	DB 12;	Length 3161;
Best Local Similarity	54.6%;	Pred. No. 0.032;		
Matches 101; Conservative	0;	Mismatches 84;	Indels 0;	Gaps 0;

[illegible]

Search completed: October 14, 1999, 04:04:20
Job time: 4558 sec

RESULT	15
MUSCDPK	
LOCUS	
DEFINITION	MUSCDPK 3161 bp mRNA ROD 11-MAY-1995
ACCESSION	Mus musculus cyclin-dependent kinase homologue (p130PITSL) mRNA complete cds.
NID	L37092
VERSION	9561745
KEYWORDS	L37092.1 GI:561745
SOURCE	cyclin-dependent protein kinase; homologue.
ORGANISM	Mus musculus cDNA to mRNA.
REFERENCE	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (Bases 1 to 3161)
JOURNAL	Malek,S.N. and Desiderio,S.
MEDLINE	A cyclin-dependent kinase homologue, p130PITSLURE is a phosphotyrosine-independent S12 ligand
FEATURES	J Biol Chem. 269 (52): 35009-35020 (1994)
source	Location/Qualifiers 1..3161

```

1..3161
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="A20"
/cell_type="lymphocyte"
662..3013
/feature="p130PITSLR"
662..3013
/feature="p130PITSLR"
662..3013
/feature="p130PITSLR"
/feature="homologue"
/codon_start=1
/product="cyclin-dependent protein kinase"
/protein_id="AA66169.1"
/db_xref="PIR:p561746"
/db_xref="GI:561746"
/translation="MGGEKQSWKVTLLDELLQEKRRKRDQEKAEIKRLKNSDRDSK
RDSLESEELDHNRKEITIRNSPYREDSDMDRDEDDSLAKPPOMSRKKAHNRKRI
EKRRKEKRHHSHSGEGKHAARVKEERERKRKRHHEDODKARMEGQKRRVAHRTY
SRKRDLELERRKEERERLEKQOEKREKREKREKREKREKREKREKREKREKREKRE
REESDQKGGHMSRLLPPRRPFMGDNKRKYKEKYEKNDLLSDLDSDSEKRT
SSAASSSESGSGSESESESESESESESESESESESESESESESESESESESESESE
OSAEVDEDEMSDEDEENHNHLLVPESESESESESESESESESESESESESESESESE
DYVDPALSPLELKEPLPYLPALGCRSVEFQOLIEEGYGVYRAKDKDTE
IVALKRIMKREKEGEPISLREINILKAGHNITVAREIVAGSMOKIYIWNVVEE
HDEKSLAEYTKQFLEGLVETITLQISGKHLHNMWLIHRDKSNLLSHACILAV
GDGFLAEYSGSPKAPATPVVLTWRAPELLIGAEYSAVDMDSVGCITFEELLQNP
LPQKSDIDINKIPDLGIPSKIMPQGINDPYAKKTFSEIPIYNNLRKFGALSD

```


1	1553.4	99.9	1555	1	T66543	Arabidopsis violana
2	624.2	40.1	1589	1	T66242	Tobacco violaxanthin
3	608.6	39.1	1981	1	T66241	Romaine lettuce vi
4	46.4	3.0	3901	1	T89345	Human p160 cDNA
5	46.4	3.0	3211	1	T89346	Human p160 cDNA
6	45.4	2.9	29332	1	V15422	Mouse poly Ig rece
7	44.4	2.9	7215	1	X13039	Enterococcus faec
8	44.2	2.8	2277	1	V05370	Human telomerase p
9	44.2	2.8	110000	1	V21209_03	Continuation (4 of)
10	43.6	2.8	2277	1	V18354	Homo sapiens ambig
11	43.2	2.8	137507	1	V19941	KSHV long unique c
12	43.2	2.8	32207	1	V73805	KSHV LTR DNA (nucl
13	42.6	2.7	1798	1	T51756	DAX-1 gene promot
14	42.2	2.7	2643	1	Q39212	CENP-B cDNA. Human
15	42	2.7	1567	1	T35277	Chemokine receptor
16	41.2	2.6	1559	1	T91855	DUB-1 enhancer/pr
17	40.6	2.6	2794	1	T65102	Luciferase gene un
18	40.2	2.6	9636	1	O67190	P. falciparum tran
19	39.9	2.6	33399	1	T05868	Chicken leucocytoz
20	39.6	2.5	53577	1	T18551	Human polycystic k
21	39.6	2.5	53577	1	T94108	Human PKD1 locus b
22	39.6	2.5	53577	1	T94108	Human PKD1 gene. H
23	39.6	2.5	53526	1	T94101	Continuation (8 of
24	39.2	2.5	110000	1	V21209_07	HSV-2 strain SB5 C
25	39.2	2.5	12700	1	V62133	g19 insert encodin
26	39	2.5	6182	1	T78867	P. falciparum liver
27	39	2.5	5361	1	T78868	P. falciparum liver
28	38.6	2.5	2010	1	N91379	Rhadinol receptor
29	38.6	2.5	15672	1	O10613	Simple tandem repe
30	38.6	2.5	494	1	Q95196	Manose-1-phosphat
31	38.4	2.5	4000	1	T91902	Genomic clone G1F
32	38.2	2.5	1505	1	O55750	Tobacco SURA-c3 mu
33	38	2.4	2520	1	O11495	Gene from the ALS
34	38	2.4	2520	1	Q28388	ALS SURA-c3 (p197E
35	38	2.4	2520	1	O81182	Tobacco herbicide-
36	38	2.4	2520	1	T33353	Continuation (5 of
37	38	2.4	110000	1	T58400_4	Tobacco C3 gene en
38	38	2.4	2520	1	T72862	LEMM1 10 cDNA. New
39	37.8	2.4	507	1	Q32658	EST clone AR354. N
40	37.8	2.4	548	1	V86412	Rat Y1321 gene. Br
41	37.4	2.4	3224	1	T91872	Continuation (14 o
42	37.4	2.4	110000	1	V21209_13	Human kidney amino
43	37.4	2.4	44453	1	X23519	

241 tcaggtcaggaattctctctaagg
 |||||'|||||
 241 tcagcttcacattctgtaaac

QY	1	caacgagtcocgagcttggatgtgtgaggaataatgagtagtgcgaagaaatgcgcaatcacac	60
Db	1	CCACGCGTCGCCCTGGTGGTGGGAGAAATTAGTAACTGTGGAAGAAATGGCAGTAGCTACAC	60
QY	61	atbgtttcaactcaactcgtgcatcgaacgcatcgaatttcgaattttccaaagtatgagtgtg	120
Db	61	ATTGTTTCACTTCACCTTGTGTCATGACCGGTATTGATTTTTCAGATGATATGATGATG	120
QY	121	gtaggcttggcgtttacagaagaagaagatgaatgaccttcttgctaaagatttacctc	180
Db	121	GTAGGCTTGGCATTACAGAGAAGAGGATCATGTGCATTTCTTGTCTCAAGATTTTACCTC	180
QY	181	caatccaaagtcyftgatactcagaacaactgtgtgagatccctcaacgctccattatcgtat	240
Db	181	CAATCCAAAGTCTGATCTCGAACAACATGTTGGAGATCCTACAGCTCTTATCTGCAT	240
QY	241	taagtgtagagttctctcaagaggaattatgaaattgtgccattaccatcaagaatgagc	300
Db	241	TAAAGTCAAGGATTCCTTAAGGGGATTTATGACATTGTGCCATTACCATCAAGAATGAGC	300

QY 301 tgaagaagctgcagctccgctgctgcttaaacctgctggtgttttaagcttgcggtcc 360
 Db 301 tgaagaagctgcagctccgctgctgcttaaacctgctggtgttttaagcttgcggtcc 360
 QY 361 ttatttttccatctgagatgcagcttgcagcttaaaccttgcagcttgcagcttaaacg 420
 Db 361 ttatttttccatctgagatgcagcttgcagcttaaaccttgcagcttgcagcttaaacg 420
 QY 421 gatccagatagaaactgcagaaatgcatgccaacccctgctgctgctgctgctgctgctg 480
 Db 421 gatccagatagaaactgcagaaatgcatgccaacccctgctgctgctgctgctgctgctg 480
 QY 481 gcttcaagctgcagcttaaacctgcagatgaaacccagctgctgctgctgctgctgctg 540
 Db 481 gcttcaagctgcagcttaaacctgcagatgaaacccagctgctgctgctgctgctgctg 540
 QY 541 tttttgagacagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 600
 Db 541 tttttgagacagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 600
 QY 601 ttccctgaaacatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 660
 Db 601 ttccctgaaacatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 660
 QY 661 acttcaacatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 720
 Db 661 acttcaacatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 720
 QY 721 ttgatgcttgcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 780
 Db 721 ttgatgcttgcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 780
 QY 781 gaaacatctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 840
 Db 781 gaaacatctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 840
 QY 841 aaaaatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 900
 Db 841 aaaaatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 900
 QY 901 ttcaacttcaagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 960
 Db 901 ttcaacttcaagatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 960
 QY 961 atatatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1020
 Db 961 atatatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1020
 QY 1021 tatacaagaagaatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1080
 Db 1021 tatacaagaagaatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1080
 QY 1081 aaagcatagagcagagcttcaagcacttcaagcagcagcagcagcagcagcagcagcagcagc 1140
 Db 1081 aaagcatagagcagagcttcaagcacttcaagcagcagcagcagcagcagcagcagcagcagc 1140
 QY 1141 ctg 1200
 Db 1141 ctg 1200
 QY 1201 aggttgaagaagatagaaagagctgaaagagctgaaagagctgaaagagctgaaagagctgaa 1260
 Db 1201 aggttgaagaagatagaaagagctgaaagagctgaaagagctgaaagagctgaaagagctgaa 1260
 QY 1261 tgaacttcttccagagatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1320
 Db 1261 tgaacttcttccagagatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1320
 QY 1321 tctgtgaagaagatagaaagagctgaaagagctgaaagagctgaaagagctgaaagagctgaa 1380
 Db 1321 tctgtgaagaagatagaaagagctgaaagagctgaaagagctgaaagagctgaaagagctgaa 1380

QY 1381 gtgaggtgaaagaaatgtttgtgggaaagctttggcaatcaggaagctcaggtgaaagaaag 1440
 Db 1381 gtgaggtgaaagaaatgtttgtgggaaagctttggcaatcaggaagctcaggtgaaagaaag 1440
 QY 1441 aaccacatctgtgttaaaactatatactacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500
 Db 1441 aaccacatctgtgttaaaactatatactacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500
 QY 1501 tttttgacagctatcatcatcattcatacaatgtgataaaaaaa 1555
 Db 1501 tttttgacagctatcatcatcattcatacaatgtgataaaaaaa 1555
 RESULT 2
 ID T66242 standard; cDNA; 1589 BP.
 AC T66242;
 DT 28-JUL-1997 (first entry)
 DE Tobacco violaxanthin de-epoxidase cDNA.
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll; lettuce; ss.
 OS Nicotiana tabacum L. cv. xanthi.
 FH Key Location/Qualifiers
 FT cds 42..1478
 FT /*tag= a
 FT translt_deptide 42..443
 FT /*tag= b
 FT mat_peptide 44..1475
 FT /*tag= c
 FT W09717447-A2.
 PD 15-MAY-1997.
 PR 07-NOV-1996; U18291.
 PR 07-NOV-1995; US-006315.
 PR 06-AUG-1996; US-023502.
 PA (CALJ) CALGENE INC.
 PI Bugos RC, Rockholm DC, Yamamoto HY;
 DR WPJ: 97-281036/25.
 DR P-PSDB: W09875.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light.
 PS Disclosure; Fig 2; 41pp; English.
 CC A cDNA clone (T66242) codes for tobacco violaxanthin de-epoxidase
 CC (VDE) (W09875), an enzyme that catalyses the de-epoxidation of
 CC violaxanthin to zeaxanthin and antheraxanthin. VDE nucleic acids
 CC (see also T66241, T66243), in sense or antisense orientation, can
 CC be used in genetic constructs, pref. also contg. a platiid
 CC translocation sequence, to modify VDE levels in plants. Increased
 CC levels result in the plant being tolerant of increased light and
 CC therefore more productive and/or more resistant to disease.
 CC low light. The photosensitivity of a range of crops, trees and
 CC ornamentals can be modified.
 SQ Sequence 1589 BP; 500 A; 261 C; 367 G; 461 T;

Query Match 40.1%; Score 624.2; DB 1; Length 1589;
 Best Local Similarity 74.3%; Pred. No. 4.8e-163;
 Matches 820; Conservative 0; Mismatches 268; Indels 15; Gaps 2;

QY 364 ttgttccatctgcagatgcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 423
 Db 424 TTGCTTCAAAAGCTGATGCGGTGATGCTCTCAAGACTTGTACTGCTTACTGAAAGAGT 483
 QY 424 gcaagatagaactgcgaagatgcctgcaacccctgctgctgctgctgctgctgctgctgctgctg 483
 Db 424 GCAAGTATGAGCTTGCAGAGTGCATTTTCGAACCTTGATGCAAGCTTAATGTGCTGCTG 543
 QY 484 ttcaagctgcaataacacgctcagatgaaacccagctgctcagatgaaatgtggtgctgtg 543
 Db 484 TCCAGACTTGCACATATAGACCTGACGAAGCAATGTCAGATATAAATGTGGATTTGT 603
 QY 544 ttggaacagctgttctgtatgagcttcaagcagctgtgctgtctgctgagaaaaaagctgtctc 603

Db 604 TTGAAAACAGTGTCTAGACAGATTGAGTTCAGTGTGAGTCTCCGAAAGAAATGCTGAC 663
 604 ctgaaacatctgatactcggagaattccctcgcccaagccctctgtctgtacgaact 663
 664 CTGGAATATCGATGTGTGCTGCTTCTGTACTGATCCAGTGTCTGTGACAGAGT 723
 664 tcaacatctcggacttaacggagagtgatcattacaagtgtgttaactcaacttt 723
 724 TTGACATGAAAGATTGTAGCGGAATGTTTCATTACTCGGGGTTTAAATCCACTTTTG 783
 724 atgacctgacgtcgcacgtcgtatgattccacagaagtgacaacaagctgttgaa 783
 784 ATGCTTTGATTGCTCAATTCATGATGTTCCATACAGAA---GAAACAAACTGTGGGGA 840
 784 acatctcttgaggaataaagaccctagacagtgatcttactagtcagcgtaacaa 843
 841 ATTATCTTGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 844 aattcgtcgaagatcctaccacactgtgttctctcaatcagacaagagtaacct 903
 901 AATTGCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 904 actatcaagatcagctgataatcctgtcatcaagaagatagagaataaactgaaact 963
 961 TCTACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 964 tattgtatactacgtgtggcgaaacagatgcttgatgtatgtgtgtgtgtat 1023
 1021 TATTGTGTACTATGAGGCGAAGGATGATGATGATGATGATGATGATGATGATGAT 1080
 1024 aacagagagatctctattaccacatcatatatacagaaactcgaagacagaca 1083
 1081 ACACAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
 1084 gcatagcagagatcctcagacatcattatagaacgagataaacaatgtgtcctga 1143
 1141 AGGTGTGGGTGATTTCAACACATTCATTAACAGACATATGATGATGATGATGATG 1200
 1144 cgtcgttgagagatctgaagagacagtgagagagtgagagagatctgttaagag 1203
 1201 CCTTTGTTGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 1204 ttgaagagatagaagagagtgagagagagagagagagagagagagagagagag 1263
 1261 TTGAGAGATGAG 1308
 1264 cctgttcagagatctgctgaagatcttaatactgaactgaagcagagagagatct 1323
 1309 CTTATTTCAGTAACCTGTTTGAAGGTTTAAAGAGCTCCACGATGAGAGAGAGCT 1368
 1324 tgaagagatgaag 1383
 1369 TAAGAGAGCTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428
 1384 aggttgaaaatctgttggaagagcttgcaatcagagagagagagagagagagag 1443
 1429 AGGAG 1488
 1444 caccatgtgtgaacataat 1466
 1489 AACATACATATATATCAT 1511

OS Lactuca sativa L. cv. romaine.
 FH Key Location/Qualifiers
 FT misc_difference 26..29
 FT /*tag= a
 FT /note= "bases 26-29 are illegible in Fig 1"
 FT misc_difference 66..72
 FT /*tag= b
 FT /note= "bases 66-72 are illegible in Fig 1"
 FT misc_difference 105..110
 FT /*tag= c
 FT /note= "bases 105-110 are illegible in Fig 1"
 FT misc_difference 147..149
 FT /*tag= d
 FT /note= "bases 147-149 are illegible in Fig 1"
 FT misc_difference 186..189
 FT /*tag= e
 FT /note= "bases 186-189 are illegible in Fig 1"
 FT misc_difference 226..227
 FT /*tag= f
 FT /note= "bases 226-227 are illegible in Fig 1"
 FT cds
 FT /*tag= g
 FT /*tag= h
 FT /*tag= i
 FT mat_peptide
 FT /*tag= j
 FT /*tag= k
 FT /*tag= l
 FT /*tag= m
 FT /*tag= n
 FT /*tag= o
 FT /*tag= p
 FT /*tag= q
 FT /*tag= r
 FT /*tag= s
 FT /*tag= t
 FT /*tag= u
 FT /*tag= v
 FT /*tag= w
 FT /*tag= x
 FT /*tag= y
 FT /*tag= z
 FT /*tag= AA
 FT /*tag= AB
 FT /*tag= AC
 FT /*tag= AD
 FT /*tag= AE
 FT /*tag= AF
 FT /*tag= AG
 FT /*tag= AH
 FT /*tag= AI
 FT /*tag= AJ
 FT /*tag= AK
 FT /*tag= AL
 FT /*tag= AM
 FT /*tag= AN
 FT /*tag= AO
 FT /*tag= AP
 FT /*tag= AQ
 FT /*tag= AR
 FT /*tag= AS
 FT /*tag= AT
 FT /*tag= AU
 FT /*tag= AV
 FT /*tag= AW
 FT /*tag= AX
 FT /*tag= AY
 FT /*tag= AZ
 FT /*tag= BA
 FT /*tag= BB
 FT /*tag= BC
 FT /*tag= BD
 FT /*tag= BE
 FT /*tag= BF
 FT /*tag= BG
 FT /*tag= BH
 FT /*tag= BI
 FT /*tag= BJ
 FT /*tag= BK
 FT /*tag= BL
 FT /*tag= BM
 FT /*tag= BN
 FT /*tag= BO
 FT /*tag= BP
 FT /*tag= BQ
 FT /*tag= BR
 FT /*tag= BS
 FT /*tag= BT
 FT /*tag= BU
 FT /*tag= BV
 FT /*tag= BW
 FT /*tag= BX
 FT /*tag= BY
 FT /*tag= BZ
 FT /*tag= CA
 FT /*tag= CB
 FT /*tag= CC
 FT /*tag= CD
 FT /*tag= CE
 FT /*tag= CF
 FT /*tag= CG
 FT /*tag= CH
 FT /*tag= CI
 FT /*tag= CJ
 FT /*tag= CK
 FT /*tag= CL
 FT /*tag= CM
 FT /*tag= CN
 FT /*tag= CO
 FT /*tag= CP
 FT /*tag= CQ
 FT /*tag= CR
 FT /*tag= CS
 FT /*tag= CT
 FT /*tag= CU
 FT /*tag= CV
 FT /*tag= CW
 FT /*tag= CX
 FT /*tag= CY
 FT /*tag= CZ
 FT /*tag= DA
 FT /*tag= DB
 FT /*tag= DC
 FT /*tag= DD
 FT /*tag= DE
 FT /*tag= DF
 FT /*tag= DG
 FT /*tag= DH
 FT /*tag= DI
 FT /*tag= DJ
 FT /*tag= DK
 FT /*tag= DL
 FT /*tag= DM
 FT /*tag= DN
 FT /*tag= DO
 FT /*tag= DP
 FT /*tag= DQ
 FT /*tag= DR
 FT /*tag= DS
 FT /*tag= DT
 FT /*tag= DU
 FT /*tag= DV
 FT /*tag= DW
 FT /*tag= DX
 FT /*tag= DY
 FT /*tag= DZ
 FT /*tag= EA
 FT /*tag= EB
 FT /*tag= EC
 FT /*tag= ED
 FT /*tag= EE
 FT /*tag= EF
 FT /*tag= EG
 FT /*tag= EH
 FT /*tag= EI
 FT /*tag= EJ
 FT /*tag= EK
 FT /*tag= EL
 FT /*tag= EM
 FT /*tag= EN
 FT /*tag= EO
 FT /*tag= EP
 FT /*tag= EQ
 FT /*tag= ER
 FT /*tag= ES
 FT /*tag= ET
 FT /*tag= EU
 FT /*tag= EV
 FT /*tag= EW
 FT /*tag= EX
 FT /*tag= EY
 FT /*tag= EZ
 FT /*tag= FA
 FT /*tag= FB
 FT /*tag= FC
 FT /*tag= FD
 FT /*tag= FE
 FT /*tag= FF
 FT /*tag= FG
 FT /*tag= FH
 FT /*tag= FI
 FT /*tag= FJ
 FT /*tag= FK
 FT /*tag= FL
 FT /*tag= FM
 FT /*tag= FN
 FT /*tag= FO
 FT /*tag= FP
 FT /*tag= FQ
 FT /*tag= FR
 FT /*tag= FS
 FT /*tag= FT
 FT /*tag= FU
 FT /*tag= FV
 FT /*tag= FW
 FT /*tag= FX
 FT /*tag= FY
 FT /*tag= FZ
 FT /*tag= GA
 FT /*tag= GB
 FT /*tag= GC
 FT /*tag= GD
 FT /*tag= GE
 FT /*tag= GF
 FT /*tag= GG
 FT /*tag= GH
 FT /*tag= GI
 FT /*tag= GJ
 FT /*tag= GK
 FT /*tag= GL
 FT /*tag= GM
 FT /*tag= GN
 FT /*tag= GO
 FT /*tag= GP
 FT /*tag= GQ
 FT /*tag= GR
 FT /*tag= GS
 FT /*tag= GT
 FT /*tag= GU
 FT /*tag= GV
 FT /*tag= GW
 FT /*tag= GX
 FT /*tag= GY
 FT /*tag= GZ
 FT /*tag= HA
 FT /*tag= HB
 FT /*tag= HC
 FT /*tag= HD
 FT /*tag= HE
 FT /*tag= HF
 FT /*tag= HG
 FT /*tag= HH
 FT /*tag= HI
 FT /*tag= HJ
 FT /*tag= HK
 FT /*tag= HL
 FT /*tag= HM
 FT /*tag= HN
 FT /*tag= HO
 FT /*tag= HP
 FT /*tag= HQ
 FT /*tag= HR
 FT /*tag= HS
 FT /*tag= HT
 FT /*tag= HU
 FT /*tag= HV
 FT /*tag= HW
 FT /*tag= HX
 FT /*tag= HY
 FT /*tag= HZ
 FT /*tag= IA
 FT /*tag= IB
 FT /*tag= IC
 FT /*tag= ID
 FT /*tag= IE
 FT /*tag= IF
 FT /*tag= IG
 FT /*tag= IH
 FT /*tag= II
 FT /*tag= IJ
 FT /*tag= IK
 FT /*tag= IL
 FT /*tag= IM
 FT /*tag= IN
 FT /*tag= IO
 FT /*tag= IP
 FT /*tag= IQ
 FT /*tag= IR
 FT /*tag= IS
 FT /*tag= IT
 FT /*tag= IU
 FT /*tag= IV
 FT /*tag= IW
 FT /*tag= IX
 FT /*tag= IY
 FT /*tag= IZ
 FT /*tag= JA
 FT /*tag= JB
 FT /*tag= JC
 FT /*tag= JD
 FT /*tag= JE
 FT /*tag= JF
 FT /*tag= JG
 FT /*tag= JH
 FT /*tag= JI
 FT /*tag= JJ
 FT /*tag= JK
 FT /*tag= JL
 FT /*tag= JM
 FT /*tag= JN
 FT /*tag= JO
 FT /*tag= JP
 FT /*tag= JQ
 FT /*tag= JR
 FT /*tag= JS
 FT /*tag= JT
 FT /*tag= JU
 FT /*tag= JV
 FT /*tag= JW
 FT /*tag= JX
 FT /*tag= JY
 FT /*tag= JZ
 FT /*tag= KA
 FT /*tag= KB
 FT /*tag= KC
 FT /*tag= KD
 FT /*tag= KE
 FT /*tag= KF
 FT /*tag= KG
 FT /*tag= KH
 FT /*tag= KI
 FT /*tag= KL
 FT /*tag= KM
 FT /*tag= KN
 FT /*tag= KO
 FT /*tag= KP
 FT /*tag= KQ
 FT /*tag= KR
 FT /*tag= KS
 FT /*tag= KT
 FT /*tag= KU
 FT /*tag= KV
 FT /*tag= KW
 FT /*tag= KX
 FT /*tag= KY
 FT /*tag= KZ
 FT /*tag= LA
 FT /*tag= LB
 FT /*tag= LC
 FT /*tag= LD
 FT /*tag= LE
 FT /*tag= LF
 FT /*tag= LG
 FT /*tag= LH
 FT /*tag= LI
 FT /*tag= LJ
 FT /*tag= LK
 FT /*tag= LL
 FT /*tag= LM
 FT /*tag= LN
 FT /*tag= LO
 FT /*tag= LP
 FT /*tag= LQ
 FT /*tag= LR
 FT /*tag= LS
 FT /*tag= LT
 FT /*tag= LU
 FT /*tag= LV
 FT /*tag= LW
 FT /*tag= LX
 FT /*tag= LY
 FT /*tag= LZ
 FT /*tag= MA
 FT /*tag= MB
 FT /*tag= MC
 FT /*tag= MD
 FT /*tag= ME
 FT /*tag= MF
 FT /*tag= MG
 FT /*tag= MH
 FT /*tag= MI
 FT /*tag= MJ
 FT /*tag= MK
 FT /*tag= ML
 FT /*tag= MN
 FT /*tag= MO
 FT /*tag= MP
 FT /*tag= MQ
 FT /*tag= MR
 FT /*tag= MS
 FT /*tag= MT
 FT /*tag= MU
 FT /*tag= MV
 FT /*tag= MW
 FT /*tag= MX
 FT /*tag= MY
 FT /*tag= MZ
 FT /*tag= NA
 FT /*tag= NB
 FT /*tag= NC
 FT /*tag= ND
 FT /*tag= NE
 FT /*tag= NF
 FT /*tag= NG
 FT /*tag= NH
 FT /*tag= NI
 FT /*tag= NJ
 FT /*tag= NK
 FT /*tag= NL
 FT /*tag= NM
 FT /*tag= NO
 FT /*tag= NP
 FT /*tag= NQ
 FT /*tag= NR
 FT /*tag= NS
 FT /*tag= NT
 FT /*tag= NU
 FT /*tag= NV
 FT /*tag= NW
 FT /*tag= NX
 FT /*tag= NY
 FT /*tag= NZ
 FT /*tag= OA
 FT /*tag= OB
 FT /*tag= OC
 FT /*tag= OD
 FT /*tag= OE
 FT /*tag= OF
 FT /*tag= OG
 FT /*tag= OH
 FT /*tag= OI
 FT /*tag= OJ
 FT /*tag= OK
 FT /*tag= OL
 FT /*tag= OM
 FT /*tag= ON
 FT /*tag= OO
 FT /*tag= OP
 FT /*tag= OQ
 FT /*tag= OR
 FT /*tag= OS
 FT /*tag= OT
 FT /*tag= OU
 FT /*tag= OV
 FT /*tag= OW
 FT /*tag= OX
 FT /*tag= OY
 FT /*tag= OZ
 FT /*tag= PA
 FT /*tag= PB
 FT /*tag= PC
 FT /*tag= PD
 FT /*tag= PE
 FT /*tag= PF
 FT /*tag= PG
 FT /*tag= PH
 FT /*tag= PI
 FT /*tag= PJ
 FT /*tag= PK
 FT /*tag= PL
 FT /*tag= PM
 FT /*tag= PN
 FT /*tag= PO
 FT /*tag= PP
 FT /*tag= PQ
 FT /*tag= PR
 FT /*tag= PS
 FT /*tag= PT
 FT /*tag= PU
 FT /*tag= PV
 FT /*tag= PW
 FT /*tag= PX
 FT /*tag= PY
 FT /*tag= PZ
 FT /*tag= QA
 FT /*tag= QB
 FT /*tag= QC
 FT /*tag= QD
 FT /*tag= QE
 FT /*tag= QF
 FT /*tag= QG
 FT /*tag= QH
 FT /*tag= QI
 FT /*tag= QJ
 FT /*tag= QK
 FT /*tag= QL
 FT /*tag= QM
 FT /*tag= QN
 FT /*tag= QO
 FT /*tag= QP
 FT /*tag= QQ
 FT /*tag= QR
 FT /*tag= QS
 FT /*tag= QT
 FT /*tag= QU
 FT /*tag= QV
 FT /*tag= QW
 FT /*tag= QX
 FT /*tag= QY
 FT /*tag= QZ
 FT /*tag= RA
 FT /*tag= RB
 FT /*tag= RC
 FT /*tag= RD
 FT /*tag= RE
 FT /*tag= RF
 FT /*tag= RG
 FT /*tag= RH
 FT /*tag= RI
 FT /*tag= RJ
 FT /*tag= RK
 FT /*tag= RL
 FT /*tag= RM
 FT /*tag= RN
 FT /*tag= RO
 FT /*tag= RP
 FT /*tag= RQ
 FT /*tag= RR
 FT /*tag= RS
 FT /*tag= RT
 FT /*tag= RU
 FT /*tag= RV
 FT /*tag= RW
 FT /*tag= RX
 FT /*tag= RY
 FT /*tag= RZ
 FT /*tag= SA
 FT /*tag= SB
 FT /*tag= SC
 FT /*tag= SD
 FT /*tag= SE
 FT /*tag= SF
 FT /*tag= SG
 FT /*tag= SH
 FT /*tag= SI
 FT /*tag= SJ
 FT /*tag= SK
 FT /*tag= SL
 FT /*tag= SM
 FT /*tag= SN
 FT /*tag= SO
 FT /*tag= SP
 FT /*tag= SQ
 FT /*tag= SR
 FT /*tag= SS
 FT /*tag= ST
 FT /*tag= SU
 FT /*tag= SV
 FT /*tag= SW
 FT /*tag= SX
 FT /*tag= SY
 FT /*tag= SZ
 FT /*tag= TA
 FT /*tag= TB
 FT /*tag= TC
 FT /*tag= TD
 FT /*tag= TE
 FT /*tag= TF
 FT /*tag= TG
 FT /*tag= TH
 FT /*tag= TI
 FT /*tag= TJ
 FT /*tag= TK
 FT /*tag= TL
 FT /*tag= TM
 FT /*tag= TN
 FT /*tag= TO
 FT /*tag= TP
 FT /*tag= TQ
 FT /*tag= TR
 FT /*tag= TS
 FT /*tag= TU
 FT /*tag= TV
 FT /*tag= TW
 FT /*tag= TX
 FT /*tag= TY
 FT /*tag= TZ
 FT /*tag= UA
 FT /*tag= UB
 FT /*tag= UC
 FT /*tag= UD
 FT /*tag= UE
 FT /*tag= UF
 FT /*tag= UG
 FT /*tag= UH
 FT /*tag= UI
 FT /*tag= UJ
 FT /*tag= UK
 FT /*tag= UL
 FT /*tag= UM
 FT /*tag= UN
 FT /*tag= UO
 FT /*tag= UP
 FT /*tag= UQ
 FT /*tag= UR
 FT /*tag= US
 FT /*tag= UT
 FT /*tag= UY
 FT /*tag= UZ
 FT /*tag= VA
 FT /*tag= VB
 FT /*tag= VC
 FT /*tag= VD
 FT /*tag= VE
 FT /*tag= VF
 FT /*tag= VG
 FT /*tag= VH
 FT /*tag= VI
 FT /*tag= VJ
 FT /*tag= VK
 FT /*tag= VL
 FT /*tag= VM
 FT /*tag= VN
 FT /*tag= VO
 FT /*tag= VP
 FT /*tag= VQ
 FT /*tag= VR
 FT /*tag= VS
 FT /*tag= VT
 FT /*tag= VY
 FT /*tag= VZ
 FT /*tag= WA
 FT /*tag= WB
 FT /*tag= WC
 FT /*tag= WD
 FT /*tag= WE
 FT /*tag= WF
 FT /*tag= WG
 FT /*tag= WH
 FT /*tag= WI
 FT /*tag= WJ
 FT /*tag= WK
 FT /*tag= WL
 FT /*tag= WM
 FT /*tag= WN
 FT /*tag= WO
 FT /*tag= WP
 FT /*tag= WQ
 FT /*tag= WR
 FT /*tag= WS
 FT /*tag= WT
 FT /*tag= WY
 FT /*tag= WZ
 FT /*tag= XA
 FT /*tag= XB
 FT /*tag= XC
 FT /*tag= XD
 FT /*tag= XE
 FT /*tag= XF
 FT /*tag= XG
 FT /*tag= XH
 FT /*tag= XI
 FT /*tag= XJ
 FT /*tag= XK
 FT /*tag= XL
 FT /*tag= XM
 FT /*tag= XN
 FT /*tag= XO
 FT /*tag= XP
 FT /*tag= XQ
 FT /*tag= XR
 FT /*tag= XS
 FT /*tag= XT
 FT /*tag= XU
 FT /*tag= XV
 FT /*tag= XW
 FT /*tag= XX
 FT /*tag= XY
 FT /*tag= XZ
 FT /*tag= YA
 FT /*tag= YB
 FT /*tag= YC
 FT /*tag= YD
 FT /*tag= YE
 FT /*tag= YF
 FT /*tag= YG
 FT /*tag= YH
 FT /*tag= YI
 FT /*tag= YJ
 FT /*tag= YK
 FT /*tag= YL
 FT /*tag= YM
 FT /*tag= YN
 FT /*tag= YO
 FT /*tag= YP
 FT /*tag= YQ
 FT /*tag= YR
 FT /*tag= YS
 FT /*tag= YT
 FT /*tag= YU
 FT /*tag= YV
 FT /*tag= YW
 FT /*tag= YX
 FT /*tag= YY
 FT /*tag= YZ
 FT /*tag= ZA
 FT /*tag= ZB
 FT /*tag= ZC
 FT /*tag= ZD
 FT /*tag= ZE
 FT /*tag= ZF
 FT /*tag= ZG
 FT /*tag= ZH
 FT /*tag= ZI
 FT /*tag= ZJ
 FT /*tag= ZK
 FT /*tag= ZL
 FT /*tag= ZM
 FT /*tag= ZN
 FT /*tag= ZO
 FT /*tag= ZP
 FT /*tag= ZQ
 FT /*tag= ZR
 FT /*tag= ZS
 FT /*tag= ZT
 FT /*tag= ZU
 FT /*tag= ZV
 FT /*tag= ZW
 FT /*tag= ZX
 FT /*tag= ZY
 FT /*tag= ZZ

Db 730 CTTGACGAGACCGAATGTCAGATTAATGCTGACTTGTGGAACAGCTGGTGAC 789
 Qy 564 gaagtcacagagtgctgtgtcgaagaaagtgtcttcctagaagaatctgatac 623
 Db 790 CAATTCAACAGAGTGTGGGTTTCCGGAAGAAATGTGCCCCGGAAATCGGATGTGGGT 849
 Qy 624 gaattctctccacagaccctctgtctgtctgtacgaacttcaaatctcgacttaac 683
 Db 850 GAATTCGCCGGTTCGGATGATAGATGAGTGGTTCACAAATTTTAATGAATGAAGACTTTAGT 909
 Qy 684 gggaaatgtgtacattcaagtggttgatccaacctgtgacccctcgactgcagctg 743
 Db 910 GGAAGAGGTATATACAGATGCTTAAATCCACATTGTGATGATGATGCAACTT 969
 Qy 744 catgagttccacacagaaagtgacacaaagcttctgtgaaataatctctgtgaaataag 803
 Db 970 CAGGAGTTTCATATGGAATATC---ATAAATTGTTGGAACTTAACATGGCGCATATAA 1026
 Qy 804 accctagacagtgatcttcttactagtcagccgtacacaaatctgtcgaagcttaac 863
 Db 1027 ACTTTGATGTGTGTTTCTTTTACTCGATCTGCTGTGCAAACTTTGTTCAAGATCCAGAT 1086
 Qy 864 caacctgtgtctctacataatcaacaaagtgatccctacatcaagatgactgtat 923
 Db 1087 CTTCCTGGAGCACTTTATATATGATGACATGATGATTTCTTCACTACCAAGATGACTGTAC 1146
 Qy 924 atctcttcaacaaagataagataaacctgagacataatctgtatactaccgtgg 983
 Db 1147 ATATTATCTTCCCAATTCGAAACAAACCCGACGATTTACATTCGTTACTACCGAGGT 1206
 Qy 984 gaaacagatgcttgagatgataatgtgtgagtgatgtatatacagaaagtgatgtat 1043
 Db 1207 CCAAAAGACGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1266
 Qy 1044 ccaaatagcattatccagaaactgaaagaaagcagaacaaacataagcagagactcagc 1103
 Db 1267 CCCGATGACCATCCCAACCTACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
 Qy 1104 acattcatgaagacgataacacatgtgtctgaacctgtcgtcgtgagagagattgag 1163
 Db 1327 AATTTCATACACACGACGATAGTGTGTGGCTGAGCTCCATTCGTGGAAGAGCTTGAG 1386
 Qy 1164 aagaaacagagaaagatgaaagatctctaaagaagtgtaagagatagaagaag 1223
 Db 1387 AAATACAGGAG 1446
 Qy 1224 gtagaagagaaagtgaaagagtgagagagtgagagagtgagagagtgagagagtgag 1283
 Db 1447 GTTGAAG 1506
 Qy 1284 gaagaaatttaataatgaactgaagcaagcagagagaaattctgtgagagagtgagaa 1343
 Db 1507 GAAGGAGTAAAG 1566
 Qy 1344 gagatgagagtggttgagagagatcaaaatgaaagcaagtgagtgagtgagtgagtgag 1403
 Db 1567 GAG 1626
 Qy 1404 aaagctttgcaactcagaaagtgagtaga 1434
 Db 1627 CGCGGTTACCGATTAGGAACCTTAGATAA 1657

RESULT 4
 T89345
 ID T89345 standard; cDNA; 3901 BP.
 AC T89345;
 DT 11-MAR-1998 (first entry)
 DE Human p160 cDNA 160.1.
 KW p160; p52; cytoplasmic; T cell; B cell; development; activation;
 KM modulation; cellular response; cell proliferation; autoimmune disease;
 OS p56-lck; ss.
 KW Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 439..3846
 FT /*tag=a
 FT /product=p160
 FT /note="160.1"
 PN MO9722255-A1.
 PE 26-JUN-1997.
 PF 11-DEC-1996; U19944.
 PR 19-DEC-1995; US-574959.
 PA (DAND) DANA FARMER CANCER INST. INC.
 PI Young I, Shin J, Strominger JL, Vedamudi RK;
 DI WPI: 97-341351/31.
 DR P-FSDB; W31185.
 PT cDNA encoding p52 and p160 and corresponding proteins - used in the
 PT treatment of autoimmune disease and for T and B cell proliferation,
 PT e.g. for treatment of tumours
 PS Claim 82; Fig 8; 175pp; English.
 CC This cDNA sequence encodes a novel p160 (160.1) which is capable
 CC of activating transcription of a variety of genes upon activation of p52
 CC and is capable of binding to the p52/p56lck complex to modulate lck
 CC function in a manner similar to p62. The genes transcribed in response to
 CC p160 activation likely include those of which are involved in T or B cell
 CC development/differentiation, T or B cell activation or production of T or
 CC B cell specific factors e.g. lymphokines or antibodies. This p160
 CC polypeptide is also a substrate for serine/threonine kinase activity.
 CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell
 CC cycle regulatory proteins stimulating expression of cell cycle dependent
 CC kinase inhibitors and arresting cell cycle progression at specific
 CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
 CC response it may be used to treat disorders where this is beneficial, e.g.
 CC infections by pathogenic microorganisms. p160 can be used to expand T
 CC cell populations for treating infectious diseases or cancer and p160
 CC inhibitors could reduce B or T cell responses and may be used to treat a
 CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,
 CC multiple sclerosis allergic reactions, Crohn's diseases etc.
 SQ Sequence 3901 BP; 781 A; 1183 C; 1128 G; 809 T;

Query Match 3.0%; Score 46.4; DB 1; Length 3901;
 Best Local Similarity 52.5%; Pred. No. 0.0041;
 Matches 126; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

Qy 1155 aagaattgagagacagtgagagagagtgagagatcgttaagagtgagagata 1214
 Db 3082 AGTATGATGAAG 3141
 Qy 1215 gaagaagagtgagagagagtgagagagtgagagagtgagagagtgagagagtgag 1274
 Db 3142 GAAG 3198
 Qy 1275 agatgtctgaagagatattatgaactcgaagcagaagcagaagaaattctgtgagagagta 1334
 Db 3199 GATGAG 3258
 Qy 1335 agtaagaagagagtgagagtgagagagtgagagagtgagagagtgagagagtgagagag 1394
 Db 3259 GAG 3318

RESULT 5
 T89346
 ID T89346 standard; cDNA; 3211 BP.
 AC T89346;
 DT 11-MAR-1998 (first entry)
 DE Human p160 cDNA 160.2.
 KW p160; p52; cytoplasmic; T cell; B cell; development; activation;
 KM modulation; cellular response; cell proliferation; autoimmune disease;
 OS p56-lck; ss.
 KW Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 439..3156
 FT /*tag=a
 FT /product=p160

FT W09122255-A1. /note= "160.2"
PN 26-JUN-1997.
PD 11-DEC-1996: U19944.
PR 19-DEC-1995: US-574959.
PA (DAND) DNA FARRER CANCER INST INC.
PI Young J, Strominger JL, Vadmudi RK;
DR WPI: 97-341351/31.
P-PDB: W1186.
PT cDNA encoding p62 and p160 and corresponding proteins - used in the
PT treatment of autoimmune disease and for T and B cell proliferation,
PT e.g. for treatment of tumours
PS Claim 82: Fig 10: 175p: English.
CC This cDNA sequence encodes a novel p160 (160.2) which is capable
CC of activating transcription of a variety of genes upon activation of p62
CC and is capable of binding to the p62/p56lck complex to modulate lck
CC function in a manner similar to p62. The genes transcribed in response to
CC p160 activation likely include those of which are involved in T or B cell
CC development/differentiation, T or B cell activation or production of T or
CC B cell specific factors e.g. lymphokines or antibodies. This p160
CC polypeptide is also a substrate for serine/threonine kinase activity.
CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell
CC cycle regulatory proteins stimulating expression of cell cycle dependent
CC kinase inhibitors and arresting cell cycle progression at specific
CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
CC response it may be used to treat disorders where this is beneficial, e.g.
CC infections by pathogenic microorganisms. p160 can be used to expand T
CC cell populations for treating infectious diseases or cancer and p160
CC inhibitors could reduce B or T cell responses and may be used to treat a
CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,
CC multiple sclerosis allergic reactions, Crohn's diseases etc.
SQ Sequence 3211 BP; 649 A; 990 C; 948 G; 624 T;

Query Match 3.0%; Score 46.4; DB 1; Length 3211;
Best Local Similarity 52.5%; Pred. No. 0.0077;
Matches 126; Conservative 0; Mismatches 111; Indels 3; Gaps 1;

QY 1155 agaatgaagaacacgtggaagtggaagatcatcgttaagagtggaagata 1214
DB 2392 AGTGATGATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2451
QY 1215 gaagaagaagtagaagaagaagtggaagtgctgtagactgagatgacctgttccag 1274
DB 2452 GAAGAAGAAGAAG 2508
QY 1275 agattgctggaagatttaataactgaagcgaagcgaagcgaagcgaagcgaagcga 1334
DB 2509 GATGAAGAGAGATATTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTT 2568
QY 1335 agtaagaagaagatgagatttgatgagatcacaatggaagcgaagtgagtggaataa 1394
DB 2569 GAGGAG 2628

RESULT 6
V15422
ID V15422 standard: DNA: 29392 BP.
AC V15422;
DT 11-JUN-1998 (first entry)
DE Mouse poly Ig receptor protein gene.
KW Mouse; poly Ig receptor protein; p1gR protein; p1g; deficiency;
KM knockout mouse; disease model; ds.
OS Mus sp.
PN 110057065-A.
PD 03-MAR-1998.
PE 19-AUG-1996; 217154.
PR 19-AUG-1996; JP-217154.
PA (HONS) YAKULT HONSHA KK.
DR WPI: 98-254323/23.
PT Mouse p1g receptor protein gene - used for preparing gene knockout
PT mice; useful for study of human poly Ig receptor protein deficiency
PS Claim 1; Page 4-14; 18pp; Japanese.

CC The present sequence represents the mouse poly Ig receptor protein
CC gene, which has a 29392 bp sequence. The new gene can be used to
CC produce a gene knockout mouse, useful as a disease model of human
CC poly Ig receptor protein deficiency.
SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

Query Match 2.9%; Score 45.4; DB 1; Length 29392;
Best Local Similarity 50.7%; Pred. No. 0.02;
Matches 109; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 1161 gagaagacagtgaagaagtggaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1220
DB 2848 GAGTGAAG 28547
QY 1221 gaggtagaagaagaagtggaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1280
DB 28548 GAGGAG 28607
QY 1281 gctgaagatttaataactgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1340
DB 28608 AAG 28667
QY 1341 gaagaagtgagcttttgatgagatcaaatgga 1375
DB 28668 AAG 28702

RESULT 7
X13039
ID X13039 standard: DNA: 7215 BP.
AC X13039;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:102.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KM vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN W09850555-A2.
PD 12-NOV-1998.
PE 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI: 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 666-669; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 7215 BP; 2685 A; 1122 C; 1520 G; 1856 T;

Query Match 2.9%; Score 44.4; DB 1; Length 7215;
Best Local Similarity 49.6%; Pred. No. 0.019;
Matches 114; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 1162 agaagacagtggaagaagtggaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1221

QY 1340 aagaagatgaggttttgatga 1363
 DB 67194 TGAAGAGTTAAAAAGATGATGA 67217

RESULT 10

AC V13834 standard; cDNA; 2277 BP.

DE Homo sapiens ambiguity-maximised telomerase protein p105 gene.
 KW telomerase; p105; treatment; prevention; cancer; restenosis;
 KW inflammation; myocardial infarction; glomerulonephritis; transplant;
 KW rejection; infection; HIV; human immunodeficiency virus;
 KW bone marrow transplants; proliferation-restricted cells;
 KW ambiguity-maximised; ds.
 OS Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT 1..2277
 FT CDS /tag=a
 FT /product= telomerase protein p105

MO9801543-A1.
 PD 15-JAN-1998.
 PF 08-JUL-1997: U12297.
 PR 08-JUL-1996: US-676967.
 PA (TUL-) TULARK INC.
 PI Cao Z;
 DR WPI; 98-101044/09.
 DR P-PSDB; W41927.
 PT New nucleic acid encoding human telomerase protein p105 or its
 PT fragments - used for therapeutic modulation of telomerase activity
 PT and for screening for potential modulators of telomerase-target
 PT binding

PS Disclosure: Pages 20-21: 32pp; English.

CC The sequence is that of an ambiguity-maximised human telomerase
 CC protein coding sequence. The sequence, or specific fragments
 CC of it, can be used to modulate expression of a telomerase

CC transcript (by hybridising to it intracellularly), e.g.

CC for treatment or prevention of cancer, restenosis, inflammation,
 CC myocardial infarction, glomerulonephritis, transplant rejection and

CC infections (e.g. with human immunodeficiency virus). It can be used to

CC express recombinant telomerase protein which can be used to screen

CC for agents, e.g. antibodies, that modulate binding of human telomerase

CC to its binding target. Those that inhibit telomerase activity can be

CC used to treat the conditions listed above, while those that are

CC agonists can be used to extend the life of proliferation-restricted

CC cells, especially normal somatic cells, e.g. in cases of

CC hypersensitivity or atrophy, also to improve production of

CC recombinant proteins by maximising cell density and survival and

CC expansion of precursor cells being used for bone marrow transplants.

CC They may also be used for diagnosis. Other uses of telomerase proteins

CC are isolation, enrichment and concentration of telomerase RNA or

CC proteins; as immunogens; in therapy; as reagent where nascent

CC oligonucleotides of known structure are needed (e.g. for tagging

CC native nucleic acid molecules) and for regulating cell growth/density

CC tolerance. The agents and the telomerase proteins should be very

CC specific, e.g. they are selective for cancer cells without harming

CC somatic cells.

CC Sequence 2277 BP: 513 A: 212 C: 395 G: 217 T:

QY 756 acagaagtgacacaagctgttggaacatctcttggaataaagaccctagacat 815

DB 226 ACNGNNGNNAARAARAYTMGNAAAYARACNARCARAARAGNARARAYGARRAAYWSN 285

QY 816 ggattcttactagtgacgcgtacaaaatctctgcaaatctcaaccactgtgtt 875

DB 286 GARTGYCCNARARAGARACNARAGCNAARARACNARAGTNGCNGAYARARARCGCMGN 345

Query Match 2.8%; Score 43.6; DB 1; Length 2277;

Best Local Similarity 25.9%; Pred. No. 0.019;

Matches 138; Conservative 86; Mismatches 308; Indels 0; Gaps 0;

QY 876 ctctacaatcatgacaacagagtaacctcaactcaatcaagatgactgtatatacctgtcatca 935
 DB 346 YTNATHATHMGNAAAYTYTNWSTTYAARTGYWSNGARGAGAYYTNARACNGTNTTYGCGN 405
 QY 936 aagatagagaat'aaacctgaagactatattgtatactaccgttggcgaaacagtgt 995
 DB 406 CARTHYYGNGCNGTNTYNGARGTNAAYATHCCNMGNAAARCCNGAYGNAARATGMGNGSN 465
 QY 996 tggatgtatattgt 1055
 DB 466 TTYGNTTYTGNCARTTYAARAAAYTYTNAGARCCNGNARACNTTNARAGNATGAY 525
 QY 1056 ataccagaactcgaataaagcagcaaaaagatagcagagacttcagacatcattaga 1115
 DB 526 ATGAARGARATTAARAGGWMGNACNGTNGCNGTNGAYTGCGCNGTNCNARAYARTAY 585
 QY 1116 acgataacacatgt 1175
 DB 586 AARGAYACNCAARWSNGTNGSNAGTNGCNGARARARARARARARARARARARARARAR 645
 QY 1176 gaagtgaaagatattgttaaagagttgaagagatagaagaagagtagaagagaa 1235
 DB 646 GARWSNGTNAARARARARAGWMGNARARARARARARARARARARARARARARARAR 705
 QY 1236 gtgagaagagtgtagtagtagtagtagtagtagtagtagtagtagtagtagtagtag 1287
 DB 706 GAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAY 757

RESULT 11

ID V19941/c
 ID V19941 standard; DNA; 137507 BP.

AC V19941;
 DT 03-AUG-1998 (first entry)

DE KSHV long unique coding region and terminal repeat.

KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;

KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;

KW complement-binding protein; glycoprotein; capsid protein IV; infection;

KW lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;

KW HIV immune status; anti-inflammatory agent; therapy; ds.

OS Kaposi's sarcoma; associated herpes virus.

FH Key Location/Qualifiers

FT 1142..2794

FT /tag=a

FT /product= complement-binding protein

FT 3699..11236

FT /tag=b

FT /product= glycoprotein B

FT complement (17261..17875)

FT /tag=c

FT /product= interleukin 6

FT complement (21548..21832)

FT /tag=d

FT /product= macrophage inflammatory protein II

FT complement (27137..27424)

FT /tag=e

FT /product= interferon regulatory factor 1

FT 28661..29741

FT /tag=f

FT /product= protein T1.1

FT complement (58976..60175)

FT /tag=g

FT /product= glycoprotein M

FT complement (69412..69915)

FT /tag=h

FT /product= glycoprotein L

FT complement (88410..88910)

FT /tag=i

FT /product= interferon regulatory factor 2

FT 89600..90541

FT /tag=j

Query Match 2.7%; Score 42.2; DB 1; Length 2643;
Best Local Similarity 48.5%; Pred. No. 0.049;
Matches 116; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 1084 gcataggcaagacttcacacatcttagaagcatacacatggtcctgaacctg 1143
DB 1148 GCTTTGGGGGTGCCCTTAATGCCACATCCACTTCCCTCAAGAGTGAAGGAGGAAG 1207
QY 1144 cgtctgtgagaaatgtgaaagacagtgaaagagtgaaagataatcgtlaaagag 1203
DB 1208 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1267
QY 1204 ttaagagataagaagaagagtgagaagaagtgagaagagtgagagtgagatga 1263
DB 1268 AAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1327
QY 1264 ccttgctcagagattgctgagagatttaagactgaagcaagagagagatttc 1322
DB 1328 TGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1386

RESULT 15

T35277/C
ID T35277 standard; CDNA; 1607 BP.
AC T35277;
DT 09-DEC-1996 (first entry)
DE Chemokine receptor K5.5 CDNA.
KW Chemokine receptor K5.5; MIP-1-alpha; RANTES; MCP-1; allergy;
KW atheroma; HIV; AIDS; graft rejection; stem cell; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 183..1265
FT /tag= a
PN WO9623068-A1.
PD 01-AUG-1996.
PF 24-JAN-1996; G00143.
PR 27-JAN-1995; GB-001683.
PA (GLAXO) GLAXO GROUP LTD.
PI Power CA, Wells TNC;
DR WPI; 96-362692/36.
DR P-PSDB; R99274.
PT Chemokine receptor which binds MIP-1-alpha, RANTES and/or MCP-1 -
PT useful in screening for agents to treat asthma, hay fever, eczema,
PT allergies, atopic dermatitis, rhinitis or conjunctivitis.
PS Example: Fig 3:47p; English.
CC A CDNA clone (T35277), designated E1-C19, codes for chemokine
CC receptor K5.5 (R99274), which binds MIP-1-alpha, RANTES and/or
CC MCP-1. It was isolated from a human spleen lambda gtl1 CDNA
CC library by screening with clone TM(2-7)5.5 (see also T35278), obtd.
CC by PCR amplification (see also T35279-80) of cDNA pred. from human
CC basophilic KU812 cells. The cDNA can be used to study gene
CC function and regulation or the design probes for chemokine
CC receptors. It may also be incorporated into a vector for prodn.
CC of recombinant K5.5 chemokine receptor, useful for screening
CC antagonist agents.
SQ Sequence 1607 BP; 355 A; 404 C; 353 G; 479 T;

Query Match 2.7%; Score 42; DB 1; Length 1607;
Best Local Similarity 66.7%; Pred. No. 0.045;
Matches 60; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1158 attgaaacacacagtgaaagagtgaaagataatcgttaagaggttgaagataaga 1217
DB 100 AATGTGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 41
QY 1218 gaagagtgagaaagaaagtgagaagtc 1247
DB 40 GAGGGGAGAAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 11

Search completed: October 14, 1999, 01:54:22
Job time: 3452 sec

KEYWORDS

EST.
thale cress.
Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
Arabidopsis.

REFERENCE

1 (bases 1 to 438)
Newman, T., deBruijn, F. J., Green, P., Keegstra, K., Kende, H.,
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,
Rezel, E. and Somerville, C.
Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)

TITLE

JOURNAL

MEDLINE

95148729
On May 8, 1995 this sequence version replaced gi:799424.

Contact: Thomas Newman

MSU-DOE Plant Research Laboratory

Michigan State University

MSU-DOE-PRU, Michigan State University, Plant Biology Bldg., E.
Lansing, MI

Tel: 517-353-0854

Fax: 517-353-9168

Email: 22313tneilm.cl.msu.edu

Seq primer: T7 dye primer.

Location/Qualifiers

1. 438

/organism="Arabidopsis thaliana"

/strain="var columbiana"

/db_xref="taxon:3702"

/clone="207C237"

/note="Vector: Lambda Zip-Lox; Site-1: Sal; Site-2: Not;

Lambda PR12 is a cDNA library derived from equal

quantities of 4 pools of mRNA. The mRNA sources were 1) 7

day germinated etiolated seedlings; 2) tissue culture

grown roots; 3) staged plants half with 24 hour light

cycle, half on 16 hr light, 8 hour dark-rose; 4)

same plants as 3 but aerial tissue (stems, flowers and

siliques. The vector is BRL's lambda Zip-Lox. The cDNA

inserts were directionally cloned with Sal-Not arms using

oligo dt primed cDNA.

BASE COUNT 108 a 97 c 102 g 114 t 17 others

ORIGIN

Query Match

Best Local Similarity 20.8%; Score 322.8; DB 24; Length 438;
Pred. No. 2.9e-72;
Matches 358; Conservative 0; Mismatches 24; Indels 5; Gaps 2;

QY 426 agaatagaactcgaagatgcatgtgcaacctgctgtgcagccaatgtcgctcctt 485
DB 14 AGGATGGAAGTCCCAAGTGCATTTGCCAACCTGCTGTGCAGCCAAATGTGCGTCTT 73
QY 486 cagaactgcaataaccgttcacagatgaacccagatgccagatgaatgtgggacatctt 545
DB 74 CAGACCTGCAATAACCGTCCACATGAACCGAGTCCACATTAAATGTGGGATCTGTTT 133
QY 546 gagaaacagtggttgatgagttcaagaagtgctgtgtgcagaaaaaagtgtgtcct 605
DB 134 GAGAAACAGTGTGTGATGATGATCAACAGAGTGTGCTGTGCGAATAAAAGTGTTCCT 123
QY 606 aaaaaatctgactcggagaatcttcctgcccagacccttctgttctgtgaagaacttc 665
DB 194 AGAAAAATCTGATCTCGGAGAAATTTCTGCCCCAGACCTTTCTGTTCTGTACAAACTTC 253
QY 666 aacatctcggacttcaacggagaagtgtacatlaaagaatgctgaatccaacttgat 725
DB 254 AACATCTCGGACTTTAAAGGGAGAGTACATTTACAAATGCTTGAAATNCAACTTTGAT 313
QY 726 gcttcagac-tgccagcgcagatgagttccacaagaagtcacaagaag-ctgttg 780
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 314 GCCTTCGNCCTGCCANCTGCATG6GTTTCNACAGAGAGTGCAACAGCTTTTGG 373
QY 781 gaacatctctggaagaataagacc 807
DB 374 AACAAACNCCTTGGGGNATAAAGCCC 400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT

2

A1668224/c

LOCUS

DEFINITION

605018D09.xl 605 - Endosperm cDNA library from Schmidt lab Zea mays

CDNA mRNA sequence.

ACCESSION

NID

VERION

KEYWORDS

SOURCE

ORGANISM

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

EST.

1

GI:4827532

BASE COUNT 149 a 164 c 117 g 172 t

ORIGIN

Query Match 10.6%; Score 164.8; DB 50; Length 602;
Best Local Similarity 66.6%; Pred. No. 5.8e-32;
Matches 257; Conservative 0; Mismatches 117; Indels 12; Gaps 1;

QY 1053 attatacagaaccgcgaagaacgcaaaagacatgagcagagacttcgacatcatt 1112
DB 596 ATAAATACCGAGGCTGGAAGAAGCTGCAAGAAGCTAGTGTGAGACTTTCGACGTTCAAC 537
QY 1113 agaaaggttaacacatgtgtctcctgaacctgcgctcgtgagagaattgagaagacagt 1172
DB 536 ATGACCGCAACACCTGTGCTGTGACCTCTCTGTGTGAGAGATCGAGAGAGGTTGAGAG 477
QY 1173 gaagaaggtgaagaagtaacatcgttaaaagaggttgtaagaatgaaagaagtagaaga 1232
DB 476 CTGGAAGAGAGAGAGACCATCTGACGAGGTGAAGAGATCGAGAGAGAGGTTGAGAG 417
QY 1233 gaagtgaagaagtcggttaagactgagatgacatttccagagattgctgaagatt 1292
DB 416 CTGGAAGAGAGG-----GAGGCGTACCTCTTTCACAAAGCTGCGAAGAGGCTTC 369
QY 1293 aatgaactgaagcagaagcaggaagaatttcgtgagagagtagaagaagaagatgag 1352
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 605018 row: D column: 09.
Location/Qualifiers
1. 602
/organism="Zea mays"
/cultivar="Oh1043"
/db_xref="taxon:4577"
/clone_id="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="pericarp, embryo, and endosperm"
/dev_stage="10 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: pAD-GAL4-2'; Site-1: EcoRI;
Site-2: XhoI; Kernel endosperm cDNA library from Schmidt
lab."

Query Match	4.58;	Score 72.2;	DB 47;	Length 584;
Best Local Similarity	78.9%;	Pred. No. 2.3e-08;		
Matches	86;	Conservative	0;	Mismatches 23; Indels 0; Gaps
Qy	364	tygttcacatcgcaatgcagftagtgtgactaaactltygcactgctttaagggat	423	
Db	476	tttccttcaaaagcctgatgccttgatgcctcaaaacttgactgcttactgaagaagt	535	
Qy	424	gcgagatagaactgcgaagaatgcattgcgaaccctgccttgcgaagcaa	472	
Db	536	gcggagataagcttgaggaaagtcactctcaaaacctgcactgacgacttga	584	

RESULT	5		
AAS50452			
LOCUS	AAS50452	621 bp	DNA
DEFINITION	1605m3 gmbpFHB3.1, G. Roman Reddy Plasmodium falciparum genomic	EST	11-AUG-1997
ACCESSION	AAS50452	clone 1605m, mRNA sequence.	

RESULT	8			
A1395403/c				
LOCUS	A1395403	390 bp	mRNA	EST
DEFINITION	MA003518.C8F Soares normalized 58W Schistosoma mansoni CDNA 3'			

sequence.
accession AA193922
nid G1783601
version AA193922.1 GI:1783601
keywords EST.
source Pristionchus pacificus.
organism Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Secernentea; Diplogasteria;

REFERENCE
AUTHORS

Diplogasterida; Diplogasterina; Diplogasteroidea
Neodiplogasterida; Pristionchus.
1 (bases 1 to 483)
Marra,M., Hiller,L., Allen,M., Bowles,M., Dietrich

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 14, 1999, 03:56:24 ; Search time 34.13 seconds
(without alignments)
979,923 Million cell updates/sec

Title: US-09-075-375-6

Perfect score: 7495

Sequence: 1 MALSLHTVLCKEALNLXA.....MEASEVEKLFGKALPIRKVR 1412

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2534	34.1	478	1	W09875	Tobacco violaxanth
2	2498	33.3	473	1	W09874	Romaine lettuce vi
3	2417	32.2	462	1	W09876	Arabidopsis violax
4	147.5	2.0	1786	1	W24790	P. falciparum live
5	147	2.0	2482	1	R72826	Human mitosis. live
6	147	2.0	2482	1	W23896	Human mitosis. amin
7	135.5	1.8	1312	1	W71295	Human homologue of
8	135.5	1.8	1312	1	W22775	Human RAD50. Human
9	135	1.8	968	1	W85011	p53alpha-green flo
10	134	1.8	1639	1	W54145	P. falciparum synt
11	132	1.8	776	1	R60178	lethal factor of B
12	129.5	1.7	2954	1	Y01632	Amino acid sequenc
13	129	1.7	1654	1	P50777	Sequence of the pl
14	129	1.7	993	1	R95268	Pre-nisin modifica
15	124	1.7	1582	1	R77087	Rat sulphonylurea
16	124	1.7	1498	1	R77084	Chicken leucocytos
17	121.5	1.6	1132	1	R97866	Nucleolar/endosoma
18	121	1.6	1411	1	W02258	TNF-R-EBA 175 fusi
19	119.5	1.6	1604	1	R70105	Mutant C-beta prot
20	119.5	1.6	1099	1	W40538	Human p160 polypep
21	119	1.6	1135	1	W31185	Human p160 polypep
22	119	1.6	905	1	W31186	Erysiphe graminis
23	118.5	1.6	2273	1	R98811	Weissia major subu
24	115.5	1.5	1130	1	R71729	Merosin major subu
25	115.5	1.5	3110	1	R71730	Merosin major subu
26	115	1.5	1435	1	R70332	Silatic acid bindin
27	115	1.5	1435	1	W22477	Potato tuber solub
28	114.5	1.5	1230	1	W17785	H. pylori GHPO 175
29	114.5	1.5	1743	1	W98879	Kineochore protel
30	112.5	1.5	3248	1	R99795	Expressed antigen
31	112	1.5	649	1	W90005	Merosin M polypept
32	111.5	1.5	1130	1	R13436	Partial BRCA2 cance
33	111.5	1.5	2329	1	W25038	Human breast and o
34	111.5	1.5	3418	1	W23287	Human BRCA2 (om14)
35	111.5	1.5	3418	1	Y04357	Human BRCA2 (om11)
36	111.5	1.5	3418	1	Y04357	Human BRCA2 (om11)
37	110.5	1.5	1786	1	R41043	CD4-BB4175 fusion
38	110.5	1.5	1164	1	R85781	Group B streptococ
39	110.5	1.5	1164	1	W40537	Group B streptococ
40	110.5	1.5	1128	1	W40539	Mutant C-beta prot
41	110.5	1.5	1093	1	W40540	Mutant C-beta prot
42	110.5	1.5	1164	1	W40541	Mutant C-beta prot
43	110	1.5	751	1	W98340	H. pylori GHPO 325

ALIGNMENTS

44	110	1.5	1392	1	Y06999	Restin protein seq
45	109.5	1.5	933	1	W98599	H. pylori GHPO 127
RESULT 1						
W09875	1					
AC	W09875	standard; Protein; 478 AA.				
DT	28-JUL-1997	(first entry)				
DE	Tobacco violaxanthin de-epoxidase.					
KW	Violaxanthin de-epoxidase; VDE; light; photosensitivity;					
KW	photoprotection; transgenic plant; zeaxanthin; antheraxanthin;					
KW	xanthophyll; tobacco.					
OS	Nicotiana tabacum cv. xanthi.					
FH	Key					
FT	peptide	Location/Qualifiers				
FT		1..134				
FT	protein	/label=Transit_peptide				
FT		135..478				
FT	peptide	/label=Mat_protein				
FT		135..147				
FT		/note="Claim 8"				
FT	domain	135..206				
FT		/label=Cys-rich_domain				
FT	domain	385..478				
FT		/label=Highly-charged_domain				
FT	misc-difference	141				
FT		/note="conserved Cys residue"				
FT	misc-difference	143				
FT		/note="conserved Cys residue"				
FT	misc-difference	148				
FT		/note="conserved Cys residue"				
FT	misc-difference	155				
FT		/note="conserved Cys residue"				
FT	misc-difference	161				
FT		/note="conserved Cys residue"				
FT	misc-difference	167				
FT		/note="conserved Cys residue"				
FT	misc-difference	171				
FT		/note="conserved Cys residue"				
FT	misc-difference	180				
FT		/note="conserved Cys residue"				
FT	misc-difference	184				
FT		/note="conserved Cys residue"				
FT	misc-difference	190				
FT		/note="conserved Cys residue"				
FT	misc-difference	206				
FT		/note="conserved Cys residue"				
FT	misc-difference	212				
FT		/note="conserved Cys residue"				
FT	misc-difference	362				
FT		/note="conserved Cys residue"				
FN	W09717447-A2.					
PN	15-MAY-1997.					
PD	07-NOV-1996; U18291.					
PR	07-NOV-1995; US-006315.					
PR	06-AUG-1996; US-923502.					
PA	(CAUL) CALGENE INC.					
PI	Bugos RC, Rockholm DC, Yamamoto HY.					
DR	WPI: 97-281036/25.					
DR	N-PDB: T66242.					
PT	DNA encoding plant violaxanthin de-epoxidase - used to modify the					
PT	sensitivity of a plant to light					
PS	Dislosure; Fig. 41pp; English.					
CC	The 55 kDa violaxanthin de-epoxidase (VDE) (W09875) of tobacco					
CC	catalyses the de-oxidation of violaxanthin to zeaxanthin and					
CC	antheraxanthin. This system, termed energy dependent non-radiative					
CC	energy dissipation, or non-photochemical fluorescence quenching,					
CC	reduces the quantum efficiency of photosystem II (PSII), helping to					
CC	prevent PSII over-reduction and photoinhibitory damage. The amino					
CC	acid sequence of the VDE was deduced from an isolated cDNA clone					

CC (T66242). VDE nucleic acids (see also T66241, T66243), in sense
 CC or antisense orientation, can be used in genetic constructs to
 CC modify VDE levels in plants. Increased levels result in the plant
 CC being tolerant of increased light and therefore more productive
 CC and/or more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 SO Sequence 478 AA:

Query Match 34.1%; Score 2554; DB 1; Length 478;
 Best Local Similarity 99.8%; Pred. No. 1,6e-193;
 Matches 476; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 474 MALAPSHNFANHEHTIKYKSLPGHKRPSWGMEDYFGSIYVAKICSSRRIPRYRKPSP 533
 DB 1 MALAPSHNFANHEHTIKYKSLPGHKRPSWGMEDYFGSIYVAKICSSRRIPRYRKPSP 60
 QY 534 RICCGDLSRGLQFSGKHNLSPAHSINQVPRKNSGCKFPKDVALLMWEKMGQFAKTAI 593
 DB 61 RICCGDLSRGLQFSGKHNLSPAHSINQVPRKNSGCKFPKDVALLMWEKMGQFAKTAI 120
 QY 594 VAIFIISVASKADAVALKTCCTCLEKCRLELAKCISNPACAAVACIQTNNRPDETEC 653
 DB 121 VAIFIISVASKADAVALKTCCTCLEKCRLELAKCISNPACAAVACIQTNNRPDETEC 180
 QY 654 QIKCGDLFENSVDNEFNECKVSRKCVPRKSDVGFVPDPSVLQKFKMKFSGKFFIT 713
 DB 181 QIKCGDLFENSVDNEFNECKVSRKCVPRKSDVGFVPDPSVLQKFKMKFSGKFFIT 240
 QY 714 RGLNPTFDAPDCOLHFEHTEENKLVGNLSWIRTPDGGFFTSRAVQKFPQDPRYPIILYN 773
 DB 241 RGLNPTFDAPDCOLHFEHTEENKLVGNLSWIRTPDGGFFTSRAVQKFPQDPRYPIILYN 300
 QY 774 HDNEHYLLYODDWTILSSKVENSPEDYIFVYKGRNDAMPDYGGSVLYTRSAVLPESTIPE 833
 DB 301 HDNEHYLLYODDWTILSSKVENSPEDYIFVYKGRNDAMPDYGGSVLYTRSAVLPESTIPE 360
 QY 834 LQTAAGKVGDEFTFRTKIDTNGEPEPLVERLEKKVEEERTLIKVEELEEEVEKVRK 893
 DB 361 LQTAAGKVGDEFTFRTKIDTNGEPEPLVERLEKKVEEERTLIKVEELEEEVEKVRK 420
 QY 894 EYTLFSKLEEGFKELQDEENFLRELKSEMDVLDGLKMEATEVEKLFGRALPIRKL 950
 DB 421 EYTLFSKLEEGFKELQDEENFLRELKSEMDVLDGLKMEATEVEKLFGRALPIRKL 477

RESULT 2

ID W09874 standard; Protein; 473 AA.
 AC W09874;
 DT 28-JUL-1997 (first entry)
 DE Romaine lettuce violaxanthin de-epoxidase.
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW Photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll; lettuce.
 OS Lactuca sativa L. cv. romaine.
 FH Key Location/Qualifiers
 FT peptide 1..125
 FT protein /label= Transl-peptide
 FT peptide 126..473
 FT peptide /label= Mat-protein
 FT peptide 126..138
 FT domain /note= "Claim 8"
 FT domain 126..197
 FT domain /label= Cys-rich domain
 FT domain 218..231
 FT domain /label= Lipocalin signature
 FT domain 376..473
 FT domain /label= Highly-charged domain
 FT peptide 265..272
 FT peptide /label= Tryptic-peptide-11
 FT peptide 275..289

FT peptide /label= Tryptic-peptide-21
 FT peptide 341..353
 FT peptide /label= Tryptic-peptide-15
 FT misc-difference 132
 FT /note= "conserved Cys residue"
 FT misc-difference 134
 FT /note= "conserved Cys residue"
 FT misc-difference 139
 FT /note= "conserved Cys residue"
 FT misc-difference 146
 FT /note= "conserved Cys residue"
 FT misc-difference 152
 FT /note= "conserved Cys residue"
 FT misc-difference 158
 FT /note= "conserved Cys residue"
 FT misc-difference 162
 FT /note= "conserved Cys residue"
 FT misc-difference 171
 FT /note= "conserved Cys residue"
 FT misc-difference 175
 FT /note= "conserved Cys residue"
 FT misc-difference 190
 FT /note= "conserved Cys residue"
 FT misc-difference 197
 FT /note= "conserved Cys residue"
 FT misc-difference 243
 FT /note= "conserved Cys residue"
 FT misc-difference 373
 FT /note= "conserved Cys residue"
 PN W09717447-A2.
 PD 15-MAY-1997.
 PE 07-NOV-1996; U18291.
 PR 07-NOV-1995; US-006315.
 PR 06-AUG-1996; US-023502.
 PA (CAL) CALGENE INC.
 PI Bugos RC, Rockholm DC, Yamamoto HY;
 DR N-PSDB: T66241.
 DR N-PSDB: T66241.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light.
 PS Example 1; Fig 1; 41pp; English.
 CC The 55 kDa violaxanthin de-epoxidase (VDE) (W09874) of romaine
 CC lettuce catalyses the de-oxidation of violaxanthin to zeaxanthin
 CC and antheraxanthin. This system, termed energy dependent
 CC non-radiative energy dissipation or non-photochemical fluorescence
 CC quenching, reduces the quantum efficiency of photosystem II (PSII)
 CC helping to prevent PSII over-reduction and photoinhibitory damage.
 CC The amino acid sequence of the VDE was deduced from an isolated
 CC cDNA clone (T66241). VDE nucleic acids (see also T66242-43), in
 CC sense or antisense orientation, can be used in genetic constructs
 CC to modify VDE levels in plants. Increased levels result in the
 CC plant being tolerant of increased light and therefore more
 CC productive and/or more resistant to disease. Underexpression of
 CC VDE increases photosynthetic efficiency under low light. The
 CC photosensitivity of a range of crops, trees and ornamentals can be
 CC modified.
 SO Sequence 473 AA:

Query Match 33.3%; Score 2498; DB 1; Length 473;
 Best Local Similarity 99.6%; Pred. No. 4,1e-189;
 Matches 471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MALSLHVFLECKEALNLYARSCNRFHRSQPPNIIIMKTRSNNGFNSFRLETSK 60
 DB 1 MALSLHVFLECKEALNLYARSCNRFHRSQPPNIIIMKTRSNNGFNSFRLETSK 60
 QY 61 TTSFSDSHCKDSQCSIDTSFEIQRFDLKGMLLILEKQROPLOAIYVCFVIV 120
 DB 61 TTSFSDSHCKDSQCSIDTSFEIQRFDLKGMLLILEKQROPLOAIYVCFVIV 120
 QY 121 PRVDADALTKACCLKECRLELAKCISNPACAAVACIQTNNRPDETECOIKCGDLFE 180

Db 121 PRYDAVDALKTACCLIKRECLACIANPSCANVACIQTCCNNRPDETCQIKGDLFE 180
 QY 181 NSVVDQFNCAVSRKKCVPRKSDVGEFPVDRNAVONFMKDFSGKWYITSGLPFTDA 240
 Db 181 NSVVDQFNCAVSRKKCVPRKSDVGEFPVDRNAVONFMKDFSGKWYITSGLPFTDA 240
 QY 241 FDCQLHEFHENDKLVGNLTWRKTLTIDGFFRISAVQITVOPDPLGALYNDNEFLHQ 300
 Db 241 FDCQLHEFHENDKLVGNLTWRKTLTIDGFFRISAVQITVOPDPLGALYNDNEFLHQ 300
 QY 301 DMYITLSSQENKPDYIFVYGRNDAMDYGGSVITRSPITLPSIIPNLQAKASVG 360
 Db 301 DMYITLSSQENKPDYIFVYGRNDAMDYGGSVITRSPITLPSIIPNLQAKASVG 360
 QY 361 RDENNFITDNSCGPEPPLVERLEKTAEGEKKLLEAVEIEBEVEKEVKYRDTMTLF 420
 Db 361 RDENNFITDNSCGPEPPLVERLEKTAEGEKKLLEAVEIEBEVEKEVKYRDTMTLF 420
 QY 421 QRLLEGFKELQDEENFVRLSEKEKILNELQMEATEVEKELGRLPIRKLK 473
 Db 421 QRLLEGFKELQDEENFVRLSEKEKILNELQMEATEVEKELGRLPIRKLK 473
 Db 421 QRLLEGFKELQDEENFVRLSEKEKILNELQMEATEVEKELGRLPIRKLK 473

RESULT 3
 W09876
 ID W09876 standard; Protein: 462 AA.
 AC W09876;
 DT 28-JUL-1997 (first entry)
 DE Arabidopsis violaxanthin de-epoxidase.
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoreaction; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll.
 OS Arabidopsis thaliana var. columbia.
 FH Key Location/Qualifiers
 FT peptide 1..113
 FT /label= Transit_peptide
 FT protein 114..462
 FT /label= Mat_Protein
 FT peptide 114..126
 FT /note= "Claim 8"
 FT domain 114..185
 FT /label= Cys-rich_domain
 FT domain 364..462
 FT /label= Highly-charged_domain
 FT misc_difference 120 /note= "conserved Cys residue"
 FT misc_difference 122 /note= "conserved Cys residue"
 FT misc_difference 127 /note= "conserved Cys residue"
 FT misc_difference 134 /note= "conserved Cys residue"
 FT misc_difference 140 /note= "conserved Cys residue"
 FT misc_difference 146 /note= "conserved Cys residue"
 FT misc_difference 150 /note= "conserved Cys residue"
 FT misc_difference 159 /note= "conserved Cys residue"
 FT misc_difference 163 /note= "conserved Cys residue"
 FT misc_difference 178 /note= "conserved Cys residue"
 FT misc_difference 185 /note= "conserved Cys residue"
 FT misc_difference 231 /note= "conserved Cys residue"
 FT misc_difference 362 /note= "conserved Cys residue"
 FT misc_difference 362 /note= "conserved Cys residue"
 PN W09717447-A2.
 PD 15-MAY-1997.
 PF 07-NOV-1996; 018291.

PR 07-NOV-1995; US-06315.
 PR 06-AUG-1996; US-023502.
 PA (CALJ) CALGENE INC.
 PI Bugos RC, RockKrohm DC, Yamamoto HY;
 DR WPI: 97-281036/25.
 DR N-PSDB: T66243.
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 PS Disclosure: Fig 3: 41pp; English.
 CC The violaxanthin de-epoxidase (VDE) (W09876) of Arabidopsis
 CC catalyzes the de-epoxidation of violaxanthin to zeaxanthin and
 CC antheraxanthin. This system, termed energy dependent non-radiative
 CC energy dissipation or non-photochemical fluorescence quenching,
 CC reduces the quantum efficiency of photosystem II (PSII), helping to
 CC prevent PSII over-reduction and photoinhibitory damage. The amino
 CC acid sequence of the VDE was deduced from an isolated cDNA clone
 CC (T66243). VDE nucleic acids (see also T66241-42), in sense or
 CC antisense orientation, can be used in genetic constructs to modify
 CC VDE levels in plants. Increased levels result in the plant being
 CC tolerant of increased light and therefore more productive and/or
 CC more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 SO Sequence 462 AA:

Query Match 32.2%; Score 2417; DB 1; Length 462;
 Best Local Similarity 99.8%; Pred. No. 9.9e-183;
 Matches 461; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 951 MAVATHCFTSPCHDRIRFFSSDDGIGRLGTRKRINGFTFLKILPIPOSADLRTGRRS 1010
 Db 1 MAVATHCFTSPCHDRIRFFSSDDGIGRLGTRKRINGFTFLKILPIPOSADLRTGRRS 60
 QY 1011 RPLSAFRSGFSKCTDIYVLPBKNELKELTAPLILKIVGACAFLYPSADAVDAKTC 1070
 Db 61 RPLSAFRSGFSKCTDIYVLPBKNELKELTAPLILKIVGACAFLYPSADAVDAKTC 120
 QY 1071 ACLIKGCRIFELAKCIANPACANVACIQTCCNNRPDETCQIKGDLFENSVDENECAV 1130
 Db 121 ACLIKGCRIFELAKCIANPACANVACIQTCCNNRPDETCQIKGDLFENSVDENECAV 180
 QY 1131 SRKCVPRKSDIGEPAPDPVSVLVONFNISDFNGKWTITSGINPFDFAFDQLHEFHREG 1190
 Db 181 SRKCVPRKSDIGEPAPDPVSVLVONFNISDFNGKWTITSGINPFDFAFDQLHEFHREG 240
 QY 1191 DNKLVGNIISMRIKTLDSGFFTRSAVQEVQDNQGVLYNHNEVLYHQDDMYILSSKIE 1250
 Db 241 DNKLVGNIISMRIKTLDSGFFTRSAVQEVQDNQGVLYNHNEVLYHQDDMYILSSKIE 300
 QY 1251 NKPEDIIFYYGRNDAMDYGGAIVYTRSSVLPNSIIPLEKKAASIGRDFSTIRTDN 1310
 Db 301 NKPEDIIFYYGRNDACDGYGAVYTRSSVLPNSIIPLEKKAASIGRDFSTIRTDN 360
 QY 1311 TCGPEPALVRIKTYVEBERIIVKVEEIEBEVEKEVKYRDTMTLFQRLAEGFNEIK 1370
 Db 361 TCGPEPALVRIKTYVEBERIIVKVEEIEBEVEKEVKYRDTMTLFQRLAEGFNEIK 420
 QY 1371 QDEENFVRLSEKEKILNELQMEASEVEKELGKALPIRKR 1412
 Db 421 QDEENFVRLSEKEKILNELQMEASEVEKELGKALPIRKR 462

RESULT 4
 W24790
 ID W24790 standard; Protein: 1786 AA.
 AC W24790;
 DT 08-OCT-1997 (first entry)
 DE P. falciparum liver stage antigen-3.
 KW Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
 KW prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;
 KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;
 KW vaccine; immunotherapy; malaria.

QY	792	-VENSPEPIFYKKGRNDADVGGSVLYTRBAVLPESI	-IPLOQTAQVGDINTF	848
Dd	1243	LEEEOE	---LNEVEDADLIKOMEKKELEKA---	LSBSDEKI 1278F
QY	849	ITDNTGCEPPPLVERLEKKEVEGE	-RTIIKEVEEI	---EEVEKYAD-----K 893
Dd	1279	IDAKD	---DTEKEVEEHDTITTLDEVELLKVVEEDKIEKVSDELDKDEIDIK	1329F
QY	894	EY----	TLFSKLPFGFKELORDEENFIRELSTKEMOVLGLMEATVEYKILFGRAL	946
Dd	1330	EYKEIKELSEIILEDYKELMT	IEDDILEEKKELEKHFKEFEDEAEIKOLEADILKVS	1389F
QY	946	-----PIKMAVAVTHCFTSPCHDRIFNFSDDGICRLGITRRKINGFFLKIL		944
Dd	1390	SLEVEEKKLEVEHELKEEVEHITISDGH	-----IKG-----	1432F
QY	995	PIIOSADLRITGRSSRPLSAPFSGSGSKITFIVPLPSKNELKEPITAPLIKIVGLACA		1054
Dd	1432	--LEEDDLEEVLDLKGSIIDMLKGMDELGMD	-----KESJEDVTTTL	1463
QY	1055	FLIVPSADAVDALKTCTCLLKGRIELACIANPACAAVACLOTCN	-NRPDETEOIK	1112
Dd	1463	-----GERVSLDVAISSALGMDDEOMKRRKAPRKLVEVYLK		1501
QY	1113	CGDLFENSVDENECASVRKCVPRKSDLGFEFAPDPSVIAONENISDFNGKWITSGL		1172
Dd	1502	-----EYKEEPEKKTITKK	-----VRFDKMD	K 1524
QY	1173	NPTFDACOLHEFHFGDKRLKIVGNISMRKITLDGFTTRSAVYOKFVODPQPOGVLYNH		1232
Dd	1525	EPRKDEIVEEMKDEDLEED	-----VEEDLEED	1552
QY	1233	NEYLHYDDWYLLSKTLENKPEDYLFVYYRGNDAMDYGAAVYVTRSSVLPNSIIPLE		1292

```

QY 1293 KAAKSGRDFSTIRDTNOCPEALVERIEKTVEEGERILKVEE-----IEEVE 1345
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1584 KRIEIV-----KAKKKLLEKKEVEGSGLEKHVDEMYKTYQKIDKEVD 1626

```

	RESULT	5	
ID	R72826		
AC	R72826 standard; Protein: 2482 AA.		
DT	27-FEB-1996 (first entry)		
DE	Human mitosis.		
KW	Cell cycle; M phase; mitosis; retinoblastome; mitosis; cell growth;		
OS	Inhibition.		
Key	Homo sapiens.		
FT	region	location/Qualifiers	
FT		1480..1659	
FT	region	/label=internal_repeat	
FT		1660..1839	
FT		/label=internal_repeat	
PN	MO9511309-A2.		
PD	27-APR-1995		
PE	24-OCT-1994; U12162.		
PR	22-OCT-1993; US-141239.		
PA	(TEXA) UNIV TEXAS SYSTEM.		
PI	Lee W, Zhu X;		
DR	WPI: 95-170229/22.		
DR	N-PDSB: Q86851.		
PT	Purified mammalian protein mitosis and agents that bind it and inhibit its action - used to promote cell growth or to inhibit cell		
PT	growth		

PF 24-OCT-1994; 328254.
 PR 24-OCT-1994; US-328254.
 PR 22-OCT-1993; US-141239.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Lee W, Zhu X;
 DR WPI; 98-109817/10.
 DR N-PSDB; V09076.
 PT New isolated mitotin protein and gene - useful for, e.g. developing
 PT products for therapy and diagnosis of hyper-proliferative disorders
 PT such as cancers or psoriasis
 PS Claim 1; Column 40-52; 43pp; English.
 CC This is the amino acid sequence for mitotin, a phosphoprotein
 CC necessary for the cell to enter mitosis. The protein's degradation is
 CC also necessary for the cell to advance into the next stages of mitosis.
 CC The mitotin antibody, can be used to control the growth of cells. An
 CC anti-mitotin antibody, a mutant or a non-functional analogue of mitotin
 CC can inhibit the mitotic cell cycle by preventing the cells from entering
 CC the M phase, and over expression of mitotin or its functional
 CC equivalent, would inhibit the cycle by preventing cells from leaving the
 CC M phase. Antagonists to this protein can be used to control
 CC hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease,
 CC psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast
 CC cancer, sarcomas and other neoplasms, bladder cancer, colon cancer,
 CC lung cancer and various leukemias and lymphomas). Reintroduction or
 CC supplementation of lost mitotin function by introduction of the protein
 CC or nucleic acid encoding the protein into a cell can restore defective
 CC chromosome segregation, which is a marker of progressing malignancy.
 CC Malignant proliferation of cells can then be halted. The protein
 CC can also be used for the detection and diagnosis of hyperproliferative
 CC cells.
 SQ Sequence 2482 AA;

Query Match 2.0%; Score 147; DB 1; Length 2482;
 Best Local Similarity 18.4%; Pred. No. 0.024;
 Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 LVELEKTAEEGKLLKEAVEIEEVE---KEVE-KVADTEMTLFFRLGFEKELQ 432
 DB 315 LSELEKTAEEGKMSILINKREIEELFOENGTLKEINASINOEKMNLOK-SSFPANYID 373
 QY 433 DEEFVVELS---KEKEEIL-----CGGLDSRGLOLFSGHK 450
 DB 374 ERESEISELDQKOEKLLLOCEETGNAVEDLSQYKAAOEKNSKLECLINECTSLGE 433
 QY 450 ---NELOMEATEVEKLEGRALP--IRKLRALAPHSNFIANHEIKYVGSKLPJG--- 501
 DB 434 NRKNELE---QLKEAFKAEHQEFLITKLAFERNOVLMLELTVQALRSEMTDONNS 489
 QY 501 KRPSMGWEDYFGSIYVAKICSSRIIPRYFRKSPRI-----CCGLDSRGLOLFSGHK 551
 DB 490 KSEGGIKOEIMTLKEONMOKEVNDLLOENOLMKVMATKHECONLESEPIR----- 544
 QY 552 HNSPAHSINONYKNGSGCKPRDYALVWEKMGQFAKAIYAIFILTSASKADAVDAL 611
 DB 544 ---NSVKEERSESRNO-CNFKPQMDLEYKESLDSYNAQVOLDAMLRNKKLEKLESE 596
 QY 612 KTCCTCL-----LKECRLEL-AKICSNP-----ACAN 637
 DB 597 KERECLQHELOTRGDELTSLNLODMQSOELSGKDEIDAEKYEISGPHLELTSQNDNAH 656
 QY 638 VAC-LQTCNNRPDETE--COIKCGDLFENSVDVEFECAVSRKKCV--PRK--SDVGF- 690
 DB 657 LQCSLOTIMKMLLEKEIEELLOAEYE--LYTELND--SRSECTIATRKMAEEVGKLL 711
 QY 690 ---PYPDSVLYOKFDMKDFSGKWFITRGILNP-----ITDAPFC 725
 DB 712 NEVKIINDSGILLAGELVEDIPGGEF---GEOPNEQHPVSLAPLDESNEYEHLLTSDKEV 768
 QY 726 QLHEPTEENKVLGNLSWRIRTPDGFFTRSAVOKFVQDPKYRGLIYNHNDNEYLLYQD-- 784
 DB 769 QMHFAELQEK-----FLSIQSHKTLIHQHQO--MSSKMSLEQLQIYVDSL 810

QY 784 ---DWITLSKVENSPEDYIFVYKGRND-----AMDGYGGSVLYTRSAVL 826
 DB 811 KAEMLVSTLNLRNFQODLVKEMQLGLEBGLVPSLSSCVPPSSLSLSSGDFY--RAL 868
 QY 827 PE---SIIEPQIAAO---KVGRDFNTFKTDITGCPPEPL-----VEFLKKEEGE 873
 DB 869 EOTGDMSTLSLNLEGAVSANQCSVDVEFCSSLOEENLIRKTPSPAPKGVLESLCEYR 928
 QY 874 RIIEKEVEIEEVEK---VRDEVTLSKLFEG-----FKELQDENEFRLSEKE 922
 DB 929 QSD---EKLEEKMSQGIKMKKEIQLBOLLSSROQLDLRKQYLSENQMOQKLTSV 984
 QY 923 EMDVDLGLKMEYVEKLP-----GRALPIRLMAVATH-----CFTS-- 961
 DB 985 TLEMESKLAEEKQTEQLSLEVARLOGLDLSRSLGIDPEDALQGRNESCDSISKE 1044
 QY 961 ---PCHDRIRFSSPDGIGRLGTRKRINGTFLKILPPI-----QSADL--- 1003
 DB 1045 HNSEETTERPKHD-VHQTCDKDAOODLMDLEKTEETGALK---PTGBCSGEOSPITNYE 1100
 QY 1003 ---RTTGRNS---SRPLAFRSFGSGIFDIYPLPSK----- 1034
 DB 1101 PGEDKTGSSSCISELSFGSPNALVPMDFLNGEDIHNLQVRKENSNTLRLLHYED 1160
 QY 1034 ---NELKETAFLLLKLVGVLAACAFILVPSADAVDALTKACCLKGRTEIACK 1084
 DB 1161 KDRKVESLNEKEKELDSKLHLQEVOLMT-----KIEAC-----IELEKI 1199
 QY 1085 IANPACAAVAVACLOTGNRPDETCQIKGDLFENSVDVEFECAVSRKKCVPRKSDLG- 1144
 DB 1200 Y-----GEL-----KR---ENSLSE 1212
 QY 1144 ---EPAPDPVLYQNFNISDFNGKWYITSGINPTFPAFCQLAEFHTE---GDNKLGVN 1197
 DB 1213 KLEFSCDHOELQVETSE-----GLNS-----DEMHAKSSREDIGNAVKN 1258
 QY 1198 ISMRIKTLDSGFEFTRSAVOKFVQDPNOGVLYNHNDNYLH---YDDWYIILSKLEK 1252
 DB 1259 DSKRKRFID-----VENLSRIIRSEKASIEHEALYLLEADLEVQVTEKLCERDNEK 1310
 QY 1253 P-----EDYIFVYRGRNDAMDGYGAVVYTRSSVLPNSIIELEKAAKSGRDEFTFI 1306
 DB 1311 QKVIYCELEELSVYTSERNQ---RGELDPMKSKTTIMDDLSMKRKKYQBELSHQSECL 1367
 QY 1307 RTDNTGCPPEPALVERIKETVEGERRIYKEVELEEB---VEKEVEKVGFTMTLFORLA 1363
 DB 1368 HClOYAEVEYKEKTELLQTLSSDVSELLKDKTHLOEKLOSLEKDSQALSTKCELENOIA 1427
 QY 1364 EGFNELKODEENFYRE-----LSKEMEFLDEIK--MEASEVEK 1400
 DB 1428 ---QLNKEKELLYKESSESLOARLSSESDYEKLNYSKALEALAEK 1468

RESULT 7
 W71295
 ID W71295 standard; Protein; 1312 AA.
 AC W71295;
 DT 25-NOV-1998 (first entry)
 DE Human homologue of yeast RAD50.
 KW Human homologue; yeast RAD50; Drosophila Septin-2; Acyl-CoA synthetase;
 OS Homo sapiens.
 PN W0938306-A1.
 PD 03-SEP-1998.
 PF 27-FEB-1997; U03159.
 PR 27-FEB-1997; WO-003159.
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Doliganov G;
 DR WPI; 98-481207/41.
 DR N-PSDB; V59979.
 PT Novel human immunomodulatory poly:peptide(s) - have homology to the
 PT yeast RAD50 or Drosophila Septin-2 proteins
 PS Disclosure; Pages 136-140; 155pp; English.

RESULT 9
 ID W85011 standard; Protein: 968 AA.
 AC W85011:
 DE 08-FEB-1999 (first entry)
 DE p85alpha-green fluorescent protein fusion product.
 KW Human, p85alpha gene; fusion protein; green fluorescent protein; GFP;
 KW Intracellular signaling; chimera.
 OS Chimera - Aequorea victoria.
 OS Chimera - Homo sapiens.
 PN WO9845704-A2.
 PD 15-OCT-1998.
 PF 07-APR-1998: DK0145.
 PR 07-APR-1997: DK-000392.
 PA (NOVO) NOVO-NORDISK AS.
 PI Kasper A, Petersen Bjorn S, Scudder K, Thastrup O,
 PI Tullin S;
 DR N-PSDB: V71026.
 PT Determining effect on signaling pathways in live cells from
 PT redistribution of luminophores - specifically fusions of green
 PT fluorescent protein with a signalling component, and new apparatus,
 PT particularly for identifying toxins and potential therapeutic agents
 PS Example 15: Pages 96-98; 326pp: English.
 CC The present sequence represents a human p85alpha-green fluorescent
 CC protein fusion product. The fusion protein is used in an assay
 CC that exemplifies the invention. The specification describes how
 CC quantitative information about the influence of a molecule on a cellular
 CC response is obtained by recording the variation, caused by the molecule,
 CC on mechanically intact living cells, in the spatially distributed light
 CC emitted from a luminophore present in the cells. The variation in light
 CC emission is processed to provide information that correlates spatial
 CC distribution to the degree of the molecule. The method is used to
 CC identify agents that (in)directly affect intracellular signalling,
 CC especially to screen for potential therapeutic agents or toxins, and
 CC to identify new drug targets.
 SQ Sequence 968 AA;

Query Match 1.8%; Score 135; DB 1: Length 968;
 Best Local Similarity 19.8%; Pred. No. 0.05;
 Matches 180; Conservative 111; Mismatches 298; Indels 318; Gaps 49;

QY 687 GDEPVPDSVL-----VOKFD-----MKDFSGKWFTRGPNRPFDAQCQHEHTEEN 735
 DB 52 GKLPYWPVLTITLYGVQCFRSYRDHMKQ-----HDFKSMAPSGYQER 97
 QY 736 KLVGMLSWRIRTPDGGFTTSAVOKFVDP-----KYPGLYNDNEYLLAYOD 783
 DB 98 TI-----EFKDDGNKYKTRAEV-KFEGDILVNRTELKGDIDFKEDGNILGKLEYNSH 149
 QY 784 DWYILSKVENSPEDEYIFVYKGRDAMDGYGSYL-----TTRSAVLPEB--IIPELQTA 837
 DB 150 NWTIADKOKNG---IVNFKIRNIED---GSYQADHYQONPIQDGVLLPD--- 199
 QY 838 AOKVGRDNTFKITDNTGPEPPLEVERLEKVEEGERTIIEVEIEEVEKVDKREYTL 897
 DB 199 -----NHLYSTQSALSKDPN-----EKRDHNVL-----LEFVTAAGITL 232
 QY 898 FSKLFEKRELOREDENLRELISKEDMD---VLDGLKKEATEVEKLGKRALPIRKMAVA 954
 DB 233 -----GNDELYK---SGIRMSAGCYORALYDKKEREDEDILHGLDITLVKGSIVA 283
 QY 955 THCFSPCHDRIRFFSSDDGIGRLGITRKRINGTFL-----LKITLPISQADLTETGGRS 1009
 DB 284 LG-FSDGGEARPEITGWLNGNETTGGEGDFFGTYVEITGKKKISPP-----TPKRP 335
 QY 1010 SRPLSAFRSGFSKGLFDI-----VPLPSKNEL---KELTAPLLKLVGV-----LACAFLL 1057
 DB 336 PRPLPLV-APGSSKTEADVEQOALTLPLDLAFOFAPPDIAPELLIKVLEIEKKGECSTLY 394
 QY 1058 -----VPSAD-----AVDAKTKACGLKGCGRF----- 1081

DB 395 RTOSSSMIAELRLDLDQDPFVSDLEMDIVHLADAFKRYLLDLPNPIPAVYSEMISLA 454
 QY 1081 -----LAKCANAPAC-----AANVACIQCTNNR----- 1104
 DB 455 PEVOSSSEYIQLLKLIRSPSPHPIQWLLQYLLKFEFLSTOTSSKNLLANAVLEISFP 514
 QY 1104 -----PDTECOIKGCDLFENSVVDFNECAVSRKRCYPRKSDLGFEFPAPPSVLY 1154
 DB 515 MLRFFSAASDNTENLIK---VIELISTEWMNRQPA-PALPCK-----PPRPTIVA 562
 QY 1155 ---QNFISDPNKNV-----ITSGLNPTFDA-----FDQQLHEHT---EGDN 1192
 DB 563 NGNNNNMSLQNAEWYGDISREEVNEKLRDADGTFILVADASTKMHGDTLLTRKGGNN 622
 QY 1193 KLVGNISWRIRKTLDS--GF--FTTSAVOKFVDP--PNQGVLYNH--DNEYLH-----QOD 1240
 DB 623 KLI-----KIFHDKGKYGSDPLTFSSVVELLNHNESLAQINRKLDKLLYPYSKIQO 677
 QY 1241 DWYILSKIE--NKPEDYIFVYKGRDAMDGYGAVVYTRSSVLPNSIPELEKAAS 1297
 DB 678 DQVVKEDNIEAVGKRLHEY-----NT---QFOEKSPRE 706
 QY 1298 IGRDFSPFIRDTNCGPEPALVERIEKV-----EEGERTIIVEEIEEIE-VEKEVE 1349
 DB 707 YDLUYERYTTSOETIMKRTALEAFNETTKIFEBOCQIOERSKEYIEFKREGNEKEIQ 766
 QY 1350 KVGRTENTLFORLAEGFNEFKODEENFVRELSEKEMEFLDEITKMASEVEKLGKALP-- 1408
 DB 767 RI-----MHNVDKRSKISIELISRRRLBEDLKKQAAEY-----RELDKMNMSIKRDL 814
 QY 1408 --IRKVR 1412:
 DB 815 IQLRKTR 821:

RESULT 10
 ID W54145
 AC W54145 standard; Protein: 1639 AA.
 DT 23-SEP-1998 (first entry)
 DE P. falciparum synthetic gp190 protein.
 KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 KW monoclonal antibody; passive immunisation; parasite.
 OS Plasmodium falciparum.
 OS Synthetic.
 PN WO9814583-A2.
 PD 09-APR-1998.
 PF 02-OCT-1997: E05441.
 PR 02-OCT-1996: DE-040817.
 PA (BUJAR) BUJAR H.
 PI Bujard H, Pan W, Tolle R;
 DR N-PSDB: V21451, V35363.
 PT Recombinant production of complete gp190/MSP-1 Plasmodium surface
 PT protein - useful in anti-malaria vaccines, also stabilising genes by
 PT reducing their AT content.
 PS Example 1: Fig 3c: 48pp: German.
 CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface) protein. The gene encoding this protein has been
 CC stabilised by reducing the AT content of the nucleotide sequence. Such a
 CC protein is useful in vaccines against malaria or for producing monoclonal
 CC antibodies (for passive immunisation). The complete gp190 protein can now
 CC be produced outside the parasite and has, at least over extended regions,
 CC the native pattern of folding. Larger amounts of the protein can be
 CC produced recombinantly than would be possible using the parasites as
 CC source.
 SQ Sequence 1639 AA;

Query Match 1.8%; Score 134; DB 1: Length 1639;
 Best Local Similarity 18.0%; Pred. No. 0.14;
 Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

```

QY 184 VQENCAVSRKRCVPRKSDVGEFVPPDRNAVONENKDESGKVIYTSGLNPTFDAEDC 243
DB 249 IENINELLEESKKTIDK-----NKNATKEEKKILYQAOYDLS-----LYNK 290
QY 244 QLEHEFMENDKLVNLTWRIKTLTDLGFEFTRSVQTF--VQDE-----DLGALYNHNE 295
DB 291 QLEEAH-----NISVLEKRIIDLTKNENIKELLDKINEIKNPPRANSNTENTLLDKRKK 346
QY 236 FLHYODDWYILSSQJENKRPDY-----IFVYGRNDAMGCGSVIYTSPLPESIT 349
DB 347 IEEHEKEIKELIKTFIDSLFTDPLLEYLYLREKKNND--ISAKVETKESTEPNE-Y 403
QY 350 PVLQAAASVSGDFNNFTTDSNCGPEPLVERLEKTAEEGKKLI-----KEAVEIEE 404
DB 404 PFGVYTPPLY-NDINNALNELNSFG--DLINPFDYTKPSKNITTDNERKKFINELKEK 459
QY 405 VEKEVEKVDTEMTLFORLEGGFKELQODEENFVELSKEKEILNEL-----QMBAT 457
DB 460 IKIEKKI-----ESDKSYEDRSKSLNDITREYKELNLEIYDSKFNNDL 507
QY 458 EYEKLEFGR--ALPIKRLMALAPHSNFIANHETIKYVGSKIPGKRKSWGMEY-PSI 514
DB 508 NEKMMGKRYSYVEKL-----THNTFASYENSKHNL-EKLTALKY--MEDISLRNI 558
QY 515 VYAKICSSRIRPRYFRKSPRICCGDLSRGLQPSHGKHNLSPAHSINONVPGNSGC-- 572
DB 559 VYKEL-----KYK-----NLSKIENFI--ELTVENIKKDEQLEK 595
QY 572 -----KEPKDVALMWEKMGOFATAIYAFILSVASKADAVDAKTCOLKECRLE- 625
DB 596 KITKDNKPKDEKILEY--SDIVKVOQKVLN-----NKIDELKKTOLIKNWELEK 645
QY 625 -----LAKCISN-----PACAA-----NVACLOICNNRPET 651
DB 646 NIHVPNSYKQENKQEBYLLIVLKEIKELKLVMPVRESLINEKKNITTEGOSDSEPT 705
QY 652 EEOI-----KCGDLFENSVDENECVSRKRCVPRKSDVGEFVPPDSVLYO-- 700
DB 706 EGEITIGATTKGQAGSALBEG--DSVQAOQOEOKQAP-----PVVPVPEKQAVP 756
QY 700 -----KFMKDFSGKWEITRGLNPTDAPDQOLHEFHTEENKLVNLSWRI 745
DB 757 TPAPVNNKTEMVSKLDYLE-----KLYEFL-----NTSYIC 788
QY 746 KRPDGFTRSAVQKVOQPKYPGILYNHD--NEYILYODDWYILSSKVEN-----SPED 798
DB 789 H-----KY--ILVSHSTNKEILKO--YKITTKEESKLSLSCDPLD 824
QY 799 YIF-----VYKGRNDAMDYGGSV-----LYTRSAVL-----PESIIPELQTAQ 839
DB 825 LLEFNIONNIPVMY--SMFDSINNSLSOLFMEIYEKENVCLYKXDKDKKNLLEEK 880
QY 840 KYGRDNFTIKTDNCTGPEP-PIVERLEKKVEGEET-----LIKVEEI----- 884
DB 881 KV-----STSVKTLSSSSMPLSLTPQDKPEVSANDTSHSTNLSNLSKLFENILSGKNK 936
QY 884 -----EEVEKYVRDKREVTLFSKLBEFKELQODEENFVELSKEEMDVLDGK 931
DB 937 NIYQELIGKSSSENYEYKILKSDTYNSESFTNFVASKADINLSLDESKR-----K 989
QY 932 MEATEVEKLEGRALPIRKILMAVAHCFSTPCDHRIFFSSDGIIGLGTTRKRNITELL 991
DB 990 LE-EDINKL-----KTLQLSFDLYNKYKLTLEKLPFKKKRYGKXMQIKKL--LL 1038
QY 992 KIILPIOSADLRTTGGSSRP--LSAFNSGSKGIFDIVPLPSKNEKLELAPELLIKLV 1048
DB 1039 K-----EOLSKSLNLPKPHVLONFSVFENK-----KKEAE----- 1071
QY 1049 GVLACAFLLYPSADAVDAKTCACLKGCGR--TELAKCIANPACANVACLOTCCNRPDE 1106
DB 1071 -----IAETENTLENTKILKHKYKGLVKKYNGESSPLKTLSEESIOTEDNYASL 1119
QY 1107 TECQIKCGDLFENSVDENECVSRKRCVPRKSDVGEFVPPDSVLYONFNISDENGKW 1166

```

```

DB 1120 E--NFKVLSLECKLKDNLN---LEKKSLSYLSSGLHHLIA-ELKEVIKKNKY----- 1167
QY 1167 YITSGNLPTDAPDQOLHEFHTEGDNKLVGNISWRKTLTDSGFTRSAVQKVO----- 1222
DB 1167 ---TGNSPS-----BNNTDVNNALF--SYK-KFLPEGTDAVATVSESGSTLEOS 1210
QY 1222 -PNOGVLVNHDEYELHYODDWYILSSKLENKPEYIFVYGRNDA--WDYGGAWVYTR 1279
DB 1211 QPKPA-----STHWGASNTITTSQNVDEVDYIIVPIGSESEYDDLGQVY--TG 1262
QY 1280 SSVLPNSIIPLEKAKSICGRDSTFIRTDNCGPEPALVERIEKTEVEGERIIVKEVEE 1339
DB 1263 EAVTPSVI-----DN-----ILSKIE--NEYEVLYLKPLAG 1291
QY 1340 IEEVEKEVEKVDTEMTLFORLAEGFNLKODEENFVELSKEEMEFID 1389
DB 1292 VYRSLEKKOLE--NNVMTFVNVKVDILNSRFKREKKNVLESLDIPYKD 1338

RESULT 11
ID R60178
AC R60178;
DT 03-APR-1995 (first entry)
DE Lethal factor of Bacillus anthracis.
KW Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KW protective antigen; cell killing; targeting; pathogen;
KW intracellular; HIV; human immunodeficiency virus; toxin.
OS Bacillus anthracis.
PN M09418332-A.
PD 18-AUG-1994.
PF 14-FEB-1994; 001524.
PR 12-FEB-1993; US-021601.
PR 23-JUN-1993; US-082849.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Acora N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
DR WPI: 94-279753/34.
DR N-PSDB: Q70179.
PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumor cells
PS or HIV-infected cells.
PS Disclosure; Page 75-77; 124pp; English.
CC The sequence encoding the lethal factor of Bacillus anthracis may be
CC used in the construction of a nucleic acid which encodes a fusion
CC protein comprising the anthrax protective antigen binding domain of
CC the native anthrax lethal factor and a sequence encoding an activity
CC inducing domain of a second protein. The fusion proteins are useful
CC for the specific killing of tumor cells or the killing of cells
CC infected with intracellular pathogens, especially HIV.
SQ Sequence 776 AA:

Query Match 1.88; Score 132; DB 1; Length 776;
Best Local Similarity 17.2%; Pred. No. 0.061;
Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

```

QY 908 ---LQRENFLELSEKEMDVLDGLKMEATEV---EKLFGRALP-----RLU 950
 Db 297 QIPEPKDDITIHLSOEKELLRQIDSSDLSTEKEFFELKQIDIRDSLSEBEKEL 356
 QY 951 MAVAHCFITSPCHDRIFRFFSSDDIGRLGTRKRINTGFLTKLIPQISAD---LRTG 1006
 Db 357 L-----NRIOVDSNP-----LSEK-EKEPLKTKLKDIOYDIDNQLQDIDG 396
 QY 1007 GRSSRLP-----SAFRSGFSKGFIDIVLPSPKNELEKLTAPPLKLKVG 1049
 Db 397 GLIDSPSINDVVRKQYKRDIONIDALHOSIGSTLYNKIYVENNMIMNLATL----- 451
 QY 1050 VLACAFIIVPSADAVALKTCACLLKGCRLTELAKCIANACANVACLOTCNNRPDETEC 1109
 Db 451 -----GADLVST----- 459
 QY 1110 QIKGDLFENSVD--FNECAVSRKCVPRK---SDLEFPAPDPSPVLVONFENISDENG 1164
 Db 459 -----DWTKINRGIFNEFKKFKYSISSNMIVDINERALDNERLKNHIOQSPDR 510
 QY 1165 KWIITSGINTFPAFCOLHEFHTBGDNKLY--GNTSMKIKTIDSGFTTRSAVQFVQDP 1222
 Db 511 AGLENG-----KLIIQRNIGLEIKDVO----- 534
 QY 1223 NQGVLYNHNDNEYLYQDDWYILSKTENKPEPYIFVYRRGRDAMDGVGAVVYTR--S 1280
 Db 534 -----IKOSKEKXIRI-DKAVYPRKSIDTKIOE---AQLNINOENKALGLPKTKLIT 584
 QY 1281 SVLPNSIIPLELEKAASIGRDEFTIRNTGCPERALVERIEKTEVGE-RIVKEV-- 1338
 Db 585 FNVHNRYSAINIVASVAILINEMKNNIQSD-----LTKVYNLYDNGSFVFDIDL 636
 QY 1338 -----EEEEVEKEVEKVGRTENTLQRLALEGNELKODEENVR----- 1380
 Db 637 PNTAEQYTHODELYEOYHSGLYVPSRSILHGPBKGV-ELRNDSGEFIHEFGHAVDY 695
 QY 1380 ---LSKEME-----FLDEIKMEASEV 1398
 Db 696 AGYLLDKNQSDLYTNKSKFIDIFKEGSLN 725

RESULT 12
 Y01632
 ID Y01632 standard; Protein: 2954 AA.
 AC Y01632;
 DE 22-JUN-1999 (first entry)
 DT Amino acid sequence of centromere-associated protein-E (CENP-E).
 KW CENP-E; centromere-associated protein-E; ATPase activity;
 KW plus end-directed microtubule motor activity; chromosome congression;
 KW microtubule binding activity; chromosome movement; mitosis;
 KW cell proliferation; tumor; metastasis; vascular malformation;
 KW inflammatory disease; immune disease; angiogenesis; hypertension;
 KW restenosis; fungal infection; selective herbicide; fungicide;
 KW insecticide; plant growth regulator; activator; cancer cell marker.
 OS Xenopus sp.
 PN WO9913061-A1.
 PD 18-MAR-1999.
 PR 10-SEP-1998; UI9231.
 PR 11-SEP-1997; US-058645.
 PA (REGC) UNIV CALIFORNIA.
 PI Cleveland DM, Goldstein LSB, Sakowicz R, Wood KW;
 DR WPL: 99-22923/19.
 DR N-PSDB: X26819.
 PR Centromere-associated protein-E and related nucleic acid
 PS Claim 5; Page 66-67; 77pp; English.
 CC The present sequence represents CENP-E (centromere-associated protein-E)
 CC of Xenopus. The protein has at least one of plus end-directed microtubule
 CC motor activity, ATPase (adenosine triphosphatase) activity and
 CC microtubule binding activity. CENP-E is the motor that powers chromosome
 CC movement toward microtubule plus ends and is essential for congression
 CC of chromosomes during mitosis. Modulators of CENP-E can thus control
 CC cell proliferation. Agents that modulate CENP-E activity are lead
 CC therapeutic, bioggricultural and diagnostic agents, e.g. for treatment

CC of unwanted cell proliferation (typical of many examples are tumors and
 CC metastases; vascular malfunction; inflammatory and immune diseases;
 CC angiogenesis; hypertension; restenosis; and fungal infections), also as
 CC plant-protection agents (selective herbicides, fungicides and
 CC insecticides) and plant growth regulators or activators for improving
 CC yields. CENP-E is also a diagnostic marker for dividing cells, including
 CC cancer cells.
 SO Sequence 2954 AA.

Query Match : 1.7%; Score 129.5; DB 1; Length 2954;
 Best Local Similarity 19.0%; Pred. No. 0.77;
 Matches 260; Conservative 190; Mismatches 441; Indels 479; Gaps 64;

QY 202 SDVEFPVPRNNAVONFNKDKSGKWIYISGLNPFDAFCQLHEFHENKLVGNLW 261
 Db 459 SDMSFPEID-DSVCTFESDFDALSMDSNGIDAEMN-----LASKVTH 502
 QY 262 RIKTLDGFFTRSAVQTFVODPDLPGALYNHNEFLHYQDDWYILSSQIENKRPDIYFY 321
 Db 503 REKT-----SLHOSMIDFGQISDSVQPHDSS-----KENQOYLPRK 540
 QY 322 YGRNDAMDGVGSGVITRSPLESTIPYLQKAASVGRDNFTTNSGCPPEPYE 381
 Db 540 -----SGDMAEKRASF-EKEITSLQOOLQSKRE-----KRELVO 574
 QY 382 RLE-KTAEDEKLLIKEA-----VEIEVEKEVEKVRTENTLF----- 421
 Db 575 SFELKIALEEQSVAKKNLEWNTNSREHSINAQVQDVEKEVVR--KEMSLDGSQYN 631
 QY 421 -----QRL-----LEGFKELQO--DEENFVRELSE-----EKEI 448
 Db 632 ASNSDLDSSVDGRKLSSSHDECIEHRKMLEQKYDLEFIENLNKSSENDKOKSSEDPF 691
 QY 449 LNELQW-EATEVEKIFGRALPIRKLMLALPHSNFLANHETIKYVYSGKLPGHKRFSWG 507
 Db 692 MESIQCEALIMAK-----ANALELALMDNPNTILLETILKRELAD----- 736
 QY 508 EDIFGSLVAKTSSRRIPRYPRKSPRICCGDLSRGLQFSHGKNLSPANSI----- 561
 Db 736 -----LERSLKENOET--NEFELLEKETQKHEBALIHEIGSLKLV 775
 QY 561 -----NONVPRKNSGCKPFXDVALMWWEKMGQFATLALFILSVASKAVALKTC 614
 Db 776 ENAEKMNQNI-----BDLETK 792
 QY 615 TDLKECELELAKCISNPACANVACLOTCNNRPDETECOIK-----CGDL 660
 Db 793 TKLLKEGEIQLAD-----LKRADVLQKKVNFPLDSVMSGSEKLCREI 836
 QY 661 FE--NSVDFENCAVSRKCVPRKSDVDGFPVPPDSVLVQFKDMKDSGKWFITRGLNP 718
 Db 837 FOLKQISDAEAVTRDAQKCESFLSE-----NLELKEKEDDS-NMYNOK--EK 883
 QY 719 TFDPAFCOLHEFTEENKLVGNLSWIRTPDGGFTTRSAVQFVODPKPGLYVHNDY 778
 Db 884 AALFLERQLETKSNKYNKKEADL-----QKEIDS-----AFMEIN-Y 919
 QY 779 LLYQDDWYILSSKVENSPEDYIFVYKGRNDAMDGVGSGVLYTRSAVLPESIIPLQTA 838
 Db 920 LN-----GLLAGKV--PRDLL-----SRVELEKKVSEFSQOL 949
 QY 839 OKYGRDPNFIKTDNCGE-----PLVRLERLKV-----EGGER--TIKEVE 881
 Db 950 EKALEKNA-LENEVCLISEYFLPNEVECLNQISKASEELIMLQKQGHSAISISROE 1008
 QY 882 EI-EEVEVEKRVKENVLFSEKLEGELODEENFLRELSEKEMDVLDGL--KMEA--- 935
 Db 1009 IIMQSESDI---LQLTDEVHTQSKVQQTLEQVL-EKKKHNDLFEKYTRNKEAEDL 1063
 QY 935 -TEVEKLFRALPIRKLMAVATHCFTSP-----CHDIRFSSDDIGRLGITRKRIIN 986
 Db 1064 LREMENIKGTMSVEYKIADTKHELEETIRQKEQLLHEKKYFFQAMQITFPI----- 1116

QY 987 GTFLIKILPIQASDLRTTGRSSRP-----LSAFRSGFKGI 1024
 Db 1116 -TPISDSLPSPSKIVE-----GNSODPIEINDYHMLALATERNNINWCLFETRSLEQV 1169
 QY 1025 FDI-VPLPS-----KNELEKELAPLLKLV-----GVLACAPLIPSA 1061
 Db 1170 IDLNTQLOSLQASIEKSDLOKPODLFEGVKLLLEKLLHGLTDSQSLIEKLOLENL 1229
 QY 1062 DAVDALKTACCLLKGCRLEAKCIANPACAAVACLOTGNNRDETCQIKCGDLFEN-- 1120
 Db 1230 EYTELKQTLQEMKNITIERNE-----LQF-NFEDLKAEHDSLOKDLSENIE 1275
 QY 1120 SVDEFNCAVSRKKCVKRSKDLGEFPAPDSVLYQNFNIDFGKWTITSGMLPTDAF 1179
 Db 1276 QSIETDELRAAOELRPOKO-----LVDSFRQOLLOCSVGIS--PNDPAV 1320
 QY 1180 DCQ-----LHEFTEGDNKLVG--NISWRIKLDSGFFTRSAVQFYODPMQ--PGVLYN 1230
 Db 1321 ANQEVNLSGEVNSLOSEMLRGERDELQSCALVSELELLAHAKVSEGENLEITKRLNG 1380
 QY 1231 HDNEXLYQDDWYILSKIEKPEDYIFVYRRGNDAMDYGAVYTRSSVLPNS--II 1288
 Db 1381 LEKELTGKSESEVLSKMLNLEKED--NNKLEQAEYSKSEKNOFLEEFSGSOQLV 1436
 QY 1289 PLEF-----KAAS-----IGRDFSTFIRDTNCGPEPALVERIEKIVEEGRIIVKVE 1338
 Db 1437 DEIEVLKAKLKAERLEIKDRDYELVOTANTLVESKLETPLOADHEE-----DSID 1490
 QY 1339 EIEEVEKEV--EKVGTMTLFORLAEFNELEKODEENFVRELSEKEME 1386
 Db 1491 RSEMEIKVLGKLEKRNQY-LLERLQEKLELSNK-----LEIQKXME 1534

RESULT 13
 ID P50777 standard; Protein: 1654 AA.
 AC P50777.
 DT 30-SEP-1991 (first entry)
 DE Sequence of the P195 protein of Plasmodium falciparum.
 KW Malaria vaccine; epitope; antigen; immunogen.
 OS Plasmodium falciparum.
 PN EP-15454-A.
 PD 11-SEP-1985.
 PE 21-FEB-1985; 301173.
 PR 22-FEB-1984; GB-004692.
 PR 26-SEP-1984; GB-024340.
 PR 21-FEB-1985; GB-004429.
 PA (WELL) WELLCOME FOUNDATION LTD.
 PI Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
 DR WPI: 85-224845/37.
 DR N-PSDB: N50530.
 PT Cloned DNA sequence encoding plasmodium falciparum protein -
 PT useful for expressing the protein for use in vaccines against
 PT malaria
 PS Claim 6, Fig 1, 5pp: English.
 CC The sequence encoding the P195 protein of Plasmodium falciparum
 CC (N50530) and a peptide comprising at least one of its epitopes
 CC (see P50777) are claimed. Also claimed is a vaccine for inducing
 CC immunity to malaria comprising the novel peptide or P195 or a
 CC peptide comprising at least one epitope when derived from the new
 CC DNA sequence, together with a carrier.
 SQ Sequence 1654 AA.

Query Match 1.7%; Score 129; DB 1; Length 1654;
 Best Local Similarity 17.5%; Pred. No. 0.34;
 Matches 238; Conservative 214; Mismatches 499; Indels 408; Gaps 67;
 QY 184 VDOFNCAVSRKKCVKRSKDLGEFPAPDSVLYQNFNIDFGKWTITSGMLPTDAF 243
 Db 249 IENINLEIESKTTIDK-----NKNATKEEEKKKLYQAOYDLS-----TYNK 290

QY 244 OLHEFHEMENDKLVGNLFWRIKTLIDGFFTRSAVQTF--VODP-----DLPGALYNHND 295
 Db 291 QLEBAH-----NLSVLEKRIIDLTKKNENIKELDKINEIKNPANSGNFPNTLLDKRKK 346
 QY 296 FLHYQDDWYILSKQLENKPDY-----IFYVYRRGNDAMDYGGSVITYTSPLPESTI 349
 Db 347 IEHEKEIKETIAKTIKFNIDSLFTDPLELEYLRKKNRID--ISAKVETESTEPNE-Y 403
 QY 350 PNIQAKAVSGVDENFNFTDNGSGPEPLVERLEKTAEEGKELI-----KEAVEIEE 404
 Db 404 PNGVTYPLSY-NDINALELNSFG---DLNPDTYEPKSNITTDERRKFTIEIKEX 459
 QY 405 VEKEVEKRDTEMTLEFQRLLEGFEKLODEENFVRELSEKEELINEL-----QMEAT 457
 Db 460 IKTEKKKI-----ESDKRSYEDRSKSLNDITKEYEKLINETIDSKFNNDIDLT 507
 QY 458 EYKELFGR--LPIRLRLAALPHSNFLNHETIKYVSKLPGKRRPSGWEDY-FQSI 514
 Db 508 NFEKMMGKRYSTYVEKL-----THPTFASYENSKHNL-EKTLKALKY--MEDYSLRNT 558
 QY 515 VVAKICSSRRIPRYEPRKSPRICCGLDLSRGLQFSGHKNLSPASHINQVPGNSGC--- 572
 Db 559 VVEKEL-----KYK-----NLISKIENI--ETLVEIKKDEQLPEK 595
 QY 572 -----KFPKDVALMWKMGQFAKTAIVAFILSVASAKADAVDAIKTCTLLKCRLE- 625
 Db 596 KITDEKRPDEKILEV--SDIVKQVQKVLIM-----NKIDELKTQTLIKKNVELKH 645
 QY 625 -----LAKCISN-----PACAA-----NVACLOTGNNRDET 651
 Db 646 NIHVPNSYKQENKOEPRYLLIVLKKLEIDKLVKMPKVESLINEKKNTIKQGSNDSEPT 705
 QY 652 ECOL-----KCGDLFENSVVDENECAYSRKKCVKRSKDVDPFPPSVLYQ-- 700
 Db 706 EGETTGATTKPQOAGSALG--DSVQAQAOEQKQAP-----PVPVYPEAKAQVP 756
 QY 700 -----KFDMDSGKWFITGRLNFTDPAFCQHEFTEENKLVGNLSMRIRTP 748
 Db 757 TPAPAVNNKTENYSKLDYLEKLY--QFLNTSYICHRYIIVSHSTNNEKILKOYK----- 809
 QY 749 DGGFTTSANQKRVQ-DPKYRGLIYNHNDNYLYQDDWYILSKVENSEDEYIFYYKGR 807
 Db 809 ----ITKEESKLSGCPPL--DLFNLQNNIPV---MSWPSLIVYHNIMYVEKE- 858
 QY 808 NDAMDYGGSYL--YTRSAVL-----PSIIPETQTAOKVGRDFNFITKTDTCGE 858
 Db 858 ----IGYVFLIMLITREKAVCNLYKKDKDKIKNLEAKKV-----STSVKTLSSSMQ 909
 QY 859 P-PLVERLEKKEEGERT-----IIEVEEI-----EEVEKY 890
 Db 910 PLSTLPQDKPEPSANDDTSHSTINLSLKLFEENILSGKKNITYOELIGKSSSENYEKEI 969
 QY 891 RDKEVTLFSKLPBEFKELQNDDEEFLRELSEKEDVDLGGKMEATEVERKFGALPIRLK 950
 Db 970 LKSDDTYNESTTFNFKVSKADDINSLNDESKR-----KLE-EDINKL-----KKT 1014
 QY 951 MAVATHCFTSCQDRIRFFSSDDGIGRLGITRKRIKNGTFLKILPIQASDLRTTGRSS 1010
 Db 1015 LQSLSDLYNKKALKLERLPDKKTKVGYKMQIKKL--LLK-----DQLESKLNSLN 1064
 QY 1011 RP--LSAFRSGSKGFLDIYPLPSKNELEKELAPLLKLVGLVACAFLLVPSADAVDAL 1067
 Db 1065 NPKHVLQNFYSVFFNK-----KKEAE-----IAEENTL 1092
 QY 1068 KTCACLLKGR--TELAKCIANPACAAVACLOTGNNRDETCQIKC--GDLFENSIV 1122
 Db 1093 EETKTLKHHYGLVAVYNGESSPLKTSSESIOTEDENYASLENKVLSKLEGLKADVLNL 1152
 QY 1123 DEFNECAVSR-----KKCVPRKSDLGEFPAPDSVLYQNFNIDFGKWTITSGLN 1173
 Db 1153 EKKKLSYLSRGLHHLIAELKEVINKKNKNTYGNPS--VNNTDV-----N 1193
 QY 1174 PTFDAFDCQLHEFTEGDNKLVGNISWRIKTL--DSGFTTSANQKRVQDDPMQGVLYNH 1231

```

Db 1194 NALST-----KFLPBGTD-----VAIVSESGSTLFSQ-----PKRA----- 1231
QY 1232 DNEYLHYODWYLLSKIKKEDYIEVYRGANDA-MDYGAVYTRSSYLPSIPE 1290
Db 1231 -STHVGAEISNTITTSQVDEVDVIVLIFGESEEDYDGLQVVT-----GEAVITSVI-- 1285
QY 1291 LERAKSISGRDSTFIRINDNTGPEPALVERIEKTEVESEKRIYAVEIEEVEKEVEK 1350
Db 1285 -----DN-----ILSKIE-----NEEYVLKPLAGVYRSLKOLE- 1316
QY 1351 VGRTEMTLFLRLEGENELKODENFVRELSEKEMEFLD 1389
Db 1316 --NNVMTFNVVWKDILNSRFNKRKNVLESIDLPIKD 1352

RESULT 14
R95268
ID R95268 standard; Protein: 993 AA.
AC R95268.
DE 01-AUG-1996 (first entry)
DE Pre-nisin modification gene nlsb product.
KM Nisin A; nlsb gene; antimicrobial; preservative; antibiotic;
OS Lactococcus lactis strain NIZO R5.
PN WO9616180-A1.
PD 30-MAY-1996.
PF 20-NOV-1995; G02699.
PR 19-NOV-1994; GB-023404.
PA (BIOT-) BIOTECHNOLOGY & BIOLOGICAL SCI RES COUNC.
PI Dodd HM, Gasson MJ.
DR WPI; 96-268616/27.
DR N-PSDB; T29660.
PT Making cell which expresses nisin but does not contain natural nlsb
PT gene - by providing cell with variant nlsb gene, and genes for nlsb
PT modification, secretion and immunity
PS Disclosure; Fig 7; 69pp; English.
CC The gene cluster nisABICPRK (see T29660 and T29661) of Lactococcus
CC lactis includes the nlsb gene coding for pre-nisin A (R95267, see
CC also R95263) and the genes for nlsb modification, secretion and
CC immunity. nlsb (R95268) and nisc (R95270) are believed to be
CC involved in reactions that modify pre-nisin; nist (R95269) is
CC similar to a transport ATPase and is involved in translocation of
CC nisin out of the cell; nist (R95271) is involved in immunity to
CC nisin. Replacement of the natural, chromosomal copy of the nlsb
CC gene with a variant nlsb gene allows produ. of high levels of nisin
CC A variants in Lactococcus lactis hosts.
SQ Sequence 993 AA:

Query Match 1.7%; Score 129; DB 1: Length 993;
Best Local Similarity 18.8%; Pred. No. 0.15;
Matches 125; Conservative 99; Mismatches 244; Indels 196; Gaps 28;

QY 297 LHYODDWYLLSSQJENKPDY-----IFYYGRANDAMDGSGSYITRSPLEDSII 349
Db 171 IKYINVIQIISECEENDYKREICETVTLQYGEYRELSEYVIGSHI-----VNHYII 224
QY 350 PNLQAKASVGRPF--NNFTITDSCGPEPPIV---ERLEKTAEGEKLILKAVELEER 404
Db 225 SNLQKDLTS---DPSNMTFLIYVAIDEDKKYIIPLKVKVFIQETSEIIEGIEKLE 281
QY 405 VEKE-----VEKVPDTEMTL---FORLLEQFK-----LQODEEN 436
Db 282 IYEMSOILENDNYIQIDILISDEINEDVYKQOLEHLAFLGNTTKSVRYTLDDYK 341
QY 437 FVEBELSEKEEILNELQMEATEVEKLFGRALPIRKLMALAPSNLANHETIKYVGSK 496
Db 342 FIEKYGVDQVQITEL-----FDSTFGIGAPYNNH---PRNDFESEPSLTYSESE 391
QY 497 LPHGRFSWGMEDYFGSIVAKICSSKRIPIRYFRKSPRICCGIDSGLOLFSHGKHLSP 556
Db 392 R-----EKYLSMYEA-----VKHNHVNINLND 413

```

```

QY 557 AHSINQVPGKNGSGCKPKDVALMWKMGQFAKTAIVAFIL-----SVASKADAVD 609
Db 414 LEHHYQKM-----DLEKKSLEQGLEFLNLAKYEKQIFLLGDVGNNGNIGASGRFS 466
QY 610 ALKTCYLLKECRLELAKCISNPACAN---VACLQCNBPDETEQIKCGDLFENS 665
Db 467 AL-----SPELTSYHRTIVDSVERENEKELTSCIEIV--LPENIR 505
QY 666 VDEFNCAVSRKKCVPRKSDVGFPPDPSSVLY-----QKFMKPFSG---KMFITGL 716
Db 506 HAVVMTSTMRKVLDFEFTSHNEVLLNTIYIGDEKEKFYARDISTOEVLKFIYSMY 565
QY 717 NPPTFADFQO-LHEPTEENKLVGNLSMRIRPPDGGFTFR----- 756
Db 566 NKTFLFNSNELRFLYEISLDK--FGNLPWELIYRDYDIPRLVDFEIVISPAKWKIWDV 623
QY 756 ---SAVQKRVQDPKYPGILY--NHDNEYLLYOD---DWYIISKVNSPEDEYIFV---Y 803
Db 624 NSKMTTRELIOSEKIEPKFEYIVGNQNKYVLSQENPLDMEILSAIKKSSKRDFTLEQ 683
QY 804 YKGRNDAMDYGGSVLYTRSAVLPEIIPETQTAQKVGRODNFPIKDNICGPEPPIVE 863
Db 684 FEDENINNGEGRV-----ADVVPFLITRA--LGNGRAPIREKRVV-----VE 727
QY 864 RLEK 867
Db 728 RREK 731

RESULT 15
R77087
ID R77087 standard; Protein: 1582 AA.
AC R77087.
DE 31-JAN-1996 (first entry)
DE Rat sulphonylurea receptor.
KM Persistent hyperinsulinaemic hypoglycaemia of infancy;
KM sulphonylurea receptor; diabetes; COS.
OS Rattus sp.
PN WO9528411-A1.
PD 26-OCT-1995.
PF 12-APR-1995; U04463.
PR 13-APR-1994; US-226972.
PR 15-MAR-1985; US-404531.
PA (BATU ) BAYLOR COLLEGE MEDICINE.
PA (TEXA ) UNIV TEXAS SYSTEM.
PI Bryan LA, Cole GU, Gage RF, Nelson DA, Thomas PM.
PI WPI; 95-373758/48.
DR N-PSDB; T01568.
PT Sulphonylurea receptors and corresp. nucleic acids - used in
PT detecting persistent hyperinsulinaemic hypoglycaemia of infancy
PS Disclosure; Page 99-104; 145pp; English.
CC Rat sulphonylurea receptor (SUR) DNA was ligated into eukaryotic
CC expression vectors containing the SV40 virus promoter and used
CC to transfect COS cells. Expression of the receptor was low
CC when SUR cDNA (T01551) was used, but high when the genomic SUR
CC sequence (T01568) was used; the product had the sequence given
CC in R77087.
SQ Sequence 1582 AA:

Query Match 1.7%; Score 124; DB 1: Length 1582;
Best Local Similarity 18.8%; Pred. No. 0.79;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

QY 868 KVEEGERTIKVEIEIEVEKVDKKEVTLFSKLEBFGK--ELQDEENPLRELSEMD 925
Db 480 KLSQAORTTL-----EYSNERLQKT---NEMLRKIKLKLXAMENIFCSVYEKTRK 528
QY 926 VLDELKMEA--TEVEKLFGRALPIRKLMAYATHCFTSPCHDIRFFSSDD---GIGRL 978
Db 529 EMTSLRAFAVYTSISIMNTAIPIAVLTIV-----GHVSFFKESDPSVAFAVL 580

```


Thu Oct 14 07:57:31 1999

us-09-075-375-6.rag

Page 13

QY 979 GTRKRINGEFL-----KILPPIQ-----SADLRTTGRSSRPLSAFRSGFSKG 1023
DB 581 SLFHLLVTPFLLSVSVSTVAKLYSVQKLSSELSAEIREOCAPREAPGQA-----G 636
QY 1024 IEDIVPL-----PSKNELKELTAPLLKLGVYLACAFLLVPSADAVDALKTACCLKG 1076
DB 637 KYQAVPLKYNRRKRPAREEVRLDGLPQ-----RLTPSTDG-DADNFCVQIIGG 684
QY 1077 C-----RIELACIANPACANVACLOTCNNRPDETCQIKGDLF 1117
DB 685 FETWPDGIPITLSNITIRIPRGOLMIVGOVGCGKSSLLLATLG-----EMOKVSGAVE 738
QY 1118 ENGVDEFNCAVSRKKCVPRKSDLGEPAPDP-----SVLVONFNI-SDF 1162
DB 739 WNSLPD-----SEGRPPQOPRAGDSGRFGCQEQRPQCYASOKPWLINATVEENITFESPF 793
QY 1163 NGRWY-----ITSGLNPTFDFACQQLHEFTEGDNKLVGNISWRIKT----- 1205
DB 794 NKQRYKVIACSLQPDIDIL-----PHGDOQIGERGINLSTGGORPDQCRPEPST 845
QY 1205 -----LDSGF-----FTRSAVOKFVODPNQPGVLYNHNDNEXLHYODDWYILSS 1247
DB 846 STEPIVFLDDPFSALDVHLSDHLMQAGILELLRDBDKRTVVLVTHKLQYLPNA-DWITAMK 904
QY 1248 KIENKPEDYIFVYVYRGNDAMDGYGAVVYTRSSVLPNSIIPLEKRAKASIGRDESTFIR 1307
DB 905 DGTIQRGTLKDFORSECQLEFHW-----KTLNMRQDQLEK-----ETVME 946
QY 1308 TDNTCGEPPALVERIEKTEVEGERIIVKEVEEIEEVEKE 1347
DB 947 RK--APEPS--QGLPRAMSSRDGLLDEDEEEEAASE 981

Search completed: October 14, 1999, 03:56:38
Job time: 2721 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 1999, 03:19:31 : Search time 33.23 Seconds
(without alignments)
1702.452 Million cell updates/sec

Title: US-09-075-375-6
Sequence: 1 MALSLHTVPLCKEALNLVY.....MEASEVEKLFGKALPIRKVR 1412

Scoring table: BLOSUM62

Searched: 122810 segs, 40065486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2557	34.1	478	2	T03750	violaxanthin de-ep
2	2430	32.4	462	2	T00708	probable violaxanth
3	155.5	2.1	886	2	H69378	putative NTPase homo
4	155.5	2.1	2829	2	A42771	reticulocyte-blind
5	155	2.1	2469	2	H36812	hypothetical prote
6	146	1.9	1935	1	A37102	myosin beta heavy
7	146	1.9	2748	2	S57976	nuclear migration
8	144.5	1.9	839	2	S54174	topoisomerase I -
9	144	1.9	2166	2	G70163	hypothetical prote
10	143.5	1.9	1558	2	B71603	myosin beta heavy
11	142.5	1.9	1935	1	S06006	myosin H3 antigen PF
12	142.5	1.9	1819	2	A71928	cag island protein
13	141	1.9	1002	2	C70319	nitrite reductase
14	140.5	1.9	1302	2	UC6009	surface-located me
15	140	1.9	800	2	F64508	hypothetical prote
16	139.5	1.9	1025	2	S54044	probable membrane
17	137.5	1.8	1875	2	S38173	myosin-like protei
18	137	1.8	852	2	B43743	probable membrane-
19	137	1.8	1676	2	E71410	probable centromer
20	137	1.8	1109	2	A40801	phosphoprotein pho
21	135	1.8	3660	1	S02041	dystrophin, muscle
22	134.5	1.8	1937	2	I38055	myosin heavy chain
23	134	1.8	1639	2	S05603	major merozoite su
24	134	1.8	978	2	A70387	conserved hypoblet
25	133.5	1.8	1934	2	T48153	beta-myosin heavy
26	133	1.8	1979	2	C71622	hypothetical prote
27	132.5	1.8	1156	2	B70356	chromosome assembl
28	132	1.8	809	1	J00032	anthrax toxin leth
29	131	1.7	1156	2	E69444	chromosome segrega
30	130.5	1.7	1631	1	SAZOK1	major merozoite su
31	130	1.7	1673	2	S48385	hypothetical prote
32	129.5	1.7	1992	2	A47297	myosin heavy chain
33	129.5	1.7	1805	2	A64224	hypothetical prote
34	129	1.7	1201	2	A35815	myosin heavy chain
35	129	1.7	1175	2	C35815	myosin heavy chain
36	129	1.7	1822	2	S44849	KL2H4.8 protein -
37	129	1.7	1179	2	F71190	probable major sur
38	128	1.7	993	2	C31915	nlsB protein - Lac
39	128	1.7	624	2	PC6003	surface membrane p

40 127.5 1.7 4910 2 S64942 probable membrane
41 127.5 1.7 1790 2 S67593 transport protein
42 126.5 1.7 2628 2 S59413 probable membrane
43 126.5 1.7 1282 2 JE0120 glycoprotein A - m
44 126 1.7 1640 2 A24594 probable major sur
45 126 1.7 993 2 C48951 nlsB protein - Lac

ALIGNMENTS

RESULT 1
T03750
violaxanthin de-epoxidase precursor - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
C:Accession: T03750
R:Bugos, R.C.; Hieber, A.D.; Yamamoto, H.Y.
J. Biol. Chem. 273, 15321-15324, 1998
A:Title: Xanthophyll cycle enzymes are members of the lipocalin family, the first ide
A:Reference number: 215054
A:Accession: T03750
A>Status: preliminary: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-478 <BUS>
A:Cross-references: EMBL:U34817, NID:G1463122, PID:G1463123
A:Experimental source: strain Xanthi; tissue-type leaf
C:Genetics:
A:Gene: TVDEL
C:Function:
A:Description: violaxanthin de-epoxidase and zeaxanthin epoxidase catalyze the additi
d in protecting the photosynthetic apparatus from excessive light
A>Note: established as member of the lipocalin family
F:1-134/Domain: transi: peptide #status predicted <TNP>
F:135-478/Product: violaxanthin de-epoxidase #status predicted <MAT>

Query Match 34.1% Score 2557, DB 2, Length 478;
Best Local Similarity 100.0% Pred. No. 9.9e-151;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 MALAPSNFLAIPETIKYVSGKLGKRFSGWMDYEGSIYVAKICSSRIIPRYFRKSP 533
|||||
D 1 MALAPSNFLAIPETIKYVSGKLGKRFSGWMDYEGSIYVAKICSSRIIPRYFRKSP 60
|||||
QY 534 RICCGDLSRGICLFSGKHNLSPAHSINONVPRKNSGCKFPDVALMWKMGQFAKTAI 593
|||||
D 61 RICCGDLSRGICLFSGKHNLSPAHSINONVPRKNSGCKFPDVALMWKMGQFAKTAI 120
|||||
QY 594 VALFILSVASKADAVDALKTCTCLKECRLELAKCISNPACAAVACLOTNNRPDETIC 653
|||||
D 121 VALFILSVASKADAVDALKTCTCLKECRLELAKCISNPACAAVACLOTNNRPDETIC 180
|||||
QY 654 QINCGDLFENSVDENNECAVSRKCVPRKSDVGDVPDPSVLQKFMKDFSGKWFIT 713
|||||
D 181 QINCGDLFENSVDENNECAVSRKCVPRKSDVGDVPDPSVLQKFMKDFSGKWFIT 240
|||||
QY 714 RGLNPFDAFCQLEHFEHTEENKLVNLSWRIPTPDGGEFTTSAYOKFVODPKRYGILYN 773
|||||
D 241 RGLNPFDAFCQLEHFEHTEENKLVNLSWRIPTPDGGEFTTSAYOKFVODPKRYGILYN 300
|||||
QY 774 HDNEYLLYQDDWYILSSKVENSPEDYIFVYKGRNDAMDGIGGSVLYTSAVLPESITPE 833
|||||
D 301 HDNEYLLYQDDWYILSSKVENSPEDYIFVYKGRNDAMDGIGGSVLYTSAVLPESITPE 360
|||||
QY 834 LQRAAQVSGRDEMTFTKTQTCGPEPPIVERLEKYEDEGERTIKVEELIEEVEVRDX 893
|||||
D 361 LQRAAQVSGRDEMTFTKTQTCGPEPPIVERLEKYEDEGERTIKVEELIEEVEVRDX 420
|||||
QY 894 EVTLFSKLEFGFKELORDEENFLRELSKEMDVLQIKMEATEVEKLFGRALPIRKTL 950
|||||
D 421 EVTLFSKLEFGFKELORDEENFLRELSKEMDVLQIKMEATEVEKLFGRALPIRKTL 477

RESULT 2

T00708

Probable violaxanthin de-epoxidase precursor - Arabidopsis thaliana

N:Alternate names: protein F22013.3

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999

C:Accession: T00708

R:Shim, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con-
eoloyis, A.; Ecker, J.R.

Submitted to the EMBL Data Library, April 1998

A:Description: Genomic sequence for Arabidopsis thaliana BAC F22013.

A:Reference number: 214200

A:Accession: T00708

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-462 <SH1>

A:Cross-references: EMBL:AC003981; NID:g3063438; PID:g3063441

C:Genetics:

A:Map position: 1

A:introns: 72/3; 128/2; 160/3; 292/2

Query Match 32.4%; Score 2430; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 6, 5e-143;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 MAVAVHCTSPCHDRIRFFSSDDGIGLIGTRKRINGTFLKILIPPIQSADLRTGGRSS 1010

Db 1 MAVAVHCTSPCHDRIRFFSSDDGIGLIGTRKRINGTFLKILIPPIQSADLRTGGRSS 60

QY 1011 RPLSAFRSGFGSGIDIVPLPSKNEKELTAPLLKLVGACAFLLVPSADAVDAKTC 1070

Db 61 RPLSAFRSGFGSGIDIVPLPSKNEKELTAPLLKLVGACAFLLVPSADAVDAKTC 120

QY 1071 ACLKGCRIELAKCIANPACANVACLOTNNRPDETECOIKCGDLENSVDFENECAY 1130

Db 121 ACLKGCRIELAKCIANPACANVACLOTNNRPDETECOIKCGDLENSVDFENECAY 180

QY 1131 SRKKCVPRKSDGEPAPDPSPVLVONENISDFNGKWTITSLNTPFAFCQHEFTTEG 1190

Db 181 SRKKCVPRKSDGEPAPDPSPVLVONENISDFNGKWTITSLNTPFAFCQHEFTTEG 240

QY 1191 DNKLIVGNISWRKILTLDSEFTFRSAVOKFVODPNQGVLYNEDNEYLHYODDWYLLSKIE 1250

Db 241 DNKLIVGNISWRKILTLDSEFTFRSAVOKFVODPNQGVLYNEDNEYLHYODDWYLLSKIE 300

QY 1251 NKPEYITVYRRNDAMDYGAVYTRSSVLPNSIIPLEKAKSIGDSEFTFRTDN 1310

Db 301 NKPEYITVYRRNDAMDYGAVYTRSSVLPNSIIPLEKAKSIGDSEFTFRTDN 360

QY 1311 TCGEPALVERIEKTEVEGERIIVKEVEEIEEVEKEVEKGFEMTLFRLAGEFNEIK 1370

Db 361 TCGEPALVERIEKTEVEGERIIVKEVEEIEEVEKEVEKGFEMTLFRLAGEFNEIK 420

QY 1371 QDEENFVRELSEKEMEFLDEIKMEASEVEKLFGRALPIRRVR 1412

Db 421 QDEENFVRELSEKEMEFLDEIKMEASEVEKLFGRALPIRRVR 462

RESULT 3

H69378

Purine NTPase homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998

C:Accession: H69378

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997A:Authors: Utechtack, T.; Cotton, M.D.; Spilg, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo-

A:Reference number: A69250; MUID:98049343

A:Accession: H69378

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-886 <KLE>

A:Cross-references: GB:AE001032; GB:AE000782; NID:g2689355; PID:g2649562; TIGR:AF1032

Query Match 2.1%; Score 155.5; DB 2; Length 886;
Best Local Similarity 20.6%; Pred. No. 0.063;
Matches 137; Conservative 85; Mismatches 221; Indels 221; Gaps 27;

QY 797 EDYIFVYKGRNDAMDYGGSV-LYTRSAVPEESIIPELOTAQKVRDEFTFKTD-N 853

Db 152 EDYIFVYKGRNDAMDYGGSV-LYTRSAVPEESIIPELOTAQKVRDEFTFKTD-N 853

QY 854 TCGEPPLVERIEKVEEERTI---KEVEIEEVEKVRDKEVTFSKLEFGFKELQR 910

Db 198 RISEEIKSIESREKLESEVRNLESRLKELEHKSRLSELRKQESSVLQEV-----R 249

QY 911 DEENFLRELSEKEMDV---LDGLKMEATEVEKLFGRALPIRRKMAVATHCTSPCHDRIR 967

Db 250 GLEEKRLERLEKQLEVERIEDLEKAKAEKEL-----KRAEY- 290

QY 968 FRSSDDGIRGLITRRKINGTFLKILPPIQSA--DLRTGGRSSRPLSAFRSGFGSGIF 1025

Db 290 -----SILEKRLSEINOALRDVKEKRGDLITREAAGIOAQLKRAE 329

QY 1026 DIVPLPSKNEKELTAPL-----LTKLVGACAFLLVPSADAVDAKTC 1070

Db 330 D-----NSKLEITKRIBELERLEKESHRLE-----TKPKMDRRQGIK- 373

QY 1071 ACLKGCRIELAKCIANPACANVACLOTNNRPDETECOIKCGDLENSVDFENECAY 1130

Db 373 -----AKLEKNLTPDKVE---KMYDLKAKAEK-KEITE 404

QY 1131 SRKKCVPRKSDGEPAPDPSPVLVONENISDFNGKWTITSLNTPFAFCQHEFTTEG 1190

Db 405 KLKLLAKKSSLTGKAOIKRAVEE-----LKSABERTCPYGRLEDEEHRKN 451

QY 1191 -----DNKLIVGNISWRKILTLDSEFTFRSAVOKFVODPNQGVLYN- 1231

Db 452 IMAEYTRKKRIAEELAADELEKLEKLEKALEKQITVAKYQMWDELALENEL 511

QY 1231 --HNDYLYHODDWYLLSKIEKNPEYITVYRRNDAMDYGAVYTRSSVLPNSII 1288

Db 512 SSHDAE-----KLSAESE-----YRKVERLDGLRGQOKILLSA---SRI 550

QY 1289 PETERAKSIGRDFSTFIRDTNTGPEPALVERIEKTEVEGERIIVKEVEEIEEVEKEV 1348

Db 551 KELKSSLEI-----EAL-KNVESEGERLHRRIRREGSELEELREV 593

QY 1349 EKVGRTEMLFQRLAEGN---ELKODENFVRELSEKEMEFLDEIKMEASEVEKLGKA 1405

Db 594 -----QSLRPFYNKWLKLEKQAESRLSESLKRR- KLEDEISEBALAKLEANGKA 641

QY 1406 LPIR 1409

Db 642 EIR 645

RESULT 4

A42771

reticulocyte-binding protein 1 - Plasmodium vivax

C:Species: Plasmodium vivax

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995

C:Accession: A42771

R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:92315338

A:Accession: A42771

A:Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2829 <GAL>

A;Experimental source: Belem strain, merozoites
A;Note: sequence extracted from NCBI backbone (

A;Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBIP:108115)

Query Match	2.1%;	Score	155.5;	DB	2;	Length	2829;
Best Local Similarity	17.4%;	Pred.	No. 0.33;				
Matches	283;	Conservative	242;	Mismatches	573;	Indels	529;
						Gaps	76;

OY	57	TSTTSFSDSHOCHKRQSQCISIDTSEFEI-----OBFDLKRGGUILLKXOMROFTOLA	110
Db	1110	TSKNELLKSXKEVEDKLNVLEONEDRYKKVAKPNEBKOLEAIRGSMXLKE-----	1161
OY	111	IVLVCTEVIYPBDAVDALKTACCLKECRIELACIANPSCAAWACLQTCNNRPDETE	170
Db	1161	-----VINKHASEMTOJESTANTLK-----SNAGORENHEJDELN-----KTK	1199
OY	171	COICGDLEF--NSVDOFNECASVRKKCVPRKSDVGEPFPDRA---VQONNM-KDF	224
Db	1200	GOMR--DIYEELKRIABELEGTAVNELKANENKANKVE-PEPERNIIGHVLERTIVERKD	1256
OY	225	SGKWITISGINPFDPADQOLHEF--HMENDKLVCNLTRITLDGGFFTRSVAQTVOXD	282
Db	1237	AGK--VVEEMSKLTKEIKLOETSDDOSONELVTYSITKHLENAG-----	1301
OY	283	PDLGALYNHDNEFLHYODDMYLSSQIENKPDDYIFYYRGRRN-DAMDGYGSVITYRS	341
Db	1301	-----YED-----VIRNEDDSIQLEKAKSULETLD-----EM	1328
OY	342	PTLPBSIIIPNIOKAK---SVGRDFNN-----FITDNS-----	373
Db	1329	KLIYOVMNLOSALOQNAGISKSELNBELGVIELLISTINYSILEYVKRNSESERYSQL	1388
OY	373	CGPEPLVERLEKTAEE--EGEKLLIREAV-----EIEEVERK-----EVEKYDT	415
Db	1389	ANGETTAEGEEKNASARLAEREL--KEQIVKDDIDSIDUKVKKIGIRREILUKMES	1446
OY	416	EWTLFORLLEGFKILOODEENFYRELSEKEKELN-----ELOME-----	456
Db	1447	ALTFFEE-SEKFKOMCSHMENNAKEGKKLEYLKNNGDGGRKANITDSOMEVGNYSKAE	1505
OY	456	-----ATEVEKLFGLALPIRLKRLALAPHSNFLANHTTYIY	492
Db	1506	HAFHTVEAOVQDKTAFCEESIAVAYTKMDLNFEESL-MKEVYKCEK-----KNDFAEKY-	1559
OY	493	VGSKUPGHKRRSMGMEDYFGSIYVAKICSSRRIPRFRRSPRICGSLDRGOJFSHOK	552
Db	1559	-SAKU-----KPYDGRl-KARYSENERKIISELKREKAV--EKKESSOLDVSTK	1603
OY	553	NLSPAHSINOVNPKGNSGCKFPKDVALLVMEVKKOGCFAKTAIVAIFLVSASKADAVALDK	612
Db	1604	SILDQDNORQOLDVSLNISGRVNKONALOYDS----ADSMSVSLPISLGAKEXKLDVK	1659
OY	613	TCOTCLKECRLELAKCISNPACAANV--ACLQTCNNRPDETBCQ-IKCGDUFENSVDZF	669
Db	1660	AA-----KESYEKENLETQVONEMSRINIVEEGSLTDIDKKITDIENDLKMKOYEEOGILOKI	1715
OY	670	NECAVSRRKKCYPR-KSDVGDPFPVPDBSVLYORKDMKDFOSCMPIIRGNLPTDARDQOLH	728
Db	1716	KENAKRKRSNFELVGESEINALLDPSIIFI-KULKLEYD-----MGDKLNKGVMKN	1766
OY	729	EFTTEENKLVCNLWSRI-----RTPDGGFFTRSVAOK	760
Db	1767	EIHGETKSYNLIETHLSNATDYSVTFPEKKQSLRELAKEEHLLRRRBEAALFLLNDLKK	1826
OY	761	FVO-----DPKPGILYNHDNEYLLYODDMYIIS-----	790
Db	1827	VESLKLKEMMKKVSAEYEGKRPHOTSYSOLYODMKTIYDELCTLNDISECSSVLNNVYS	1886
OY	790	--SKVENSGE-YUITVVYKGRNDAMDGOGGSVLYLTRSAVLPRESIIPELTOAOKGRPFN	846
Db	1887	IYAKVESKHADY-----FRDANSMTESKAVTLANTFLDSEKAKISS-----GMEBN	1931

QY	847	TFIKTD-----	NTGPEPPIVERLE-----	KKVEBGR-----	TIKEK	880
		:	:	:	:	
Db	1932	AEMKSNFKTDLELEIFSVISNSNELIKRLEODSDNDVIOKERESEOLAKADTDIYNVTKL				1191
QY	881	EIEIEEVEKVRCEVFLFSKLEGEFEKLORDE-----	ENFLRELSKEENDVLGLKMEA			934
		:	:	:	:	
Db	1992	NEFNEKLEEKAKNEEYSEKVRALKRLSQVGCINGCHENHRLLDNTE				2041
QY	935	TEVEKLEGRALPIRLMAVAATHCFTSPCHDRIRFPSSDDGJGRGLITRKIRINGIFELKIL				994
		:	:	:	:	
Db	2041	-ELEBNL-----	KKWATYI-----	PDKKSERESGLOEM-----	EENEXN-TYNSNI-	2078
QY	995	PPIOSADLRITGGRSSRPILSAFNSGFSKQIPVILVPSKNE-----	LKELTAPLLKLGV			1050
Db	2078	-TDEGIVVSGBSKE-----	DIEKLESNEMERNISEKISTIDSK----			2118
QY	1051	LACAFILVPSADAVDAIKTCACILKGCRIELAKCANPACANVACLOTCNNRPDETECQ				1110
Db	2118	-----	VIENKSTIDELIK--	LGRMCQAHMISLISYANKTKSKKLIMKEKENE--		2167
QY	1111	IKCGDLEF--	NSVDFENECANVRKKCVPRKSDLG--	EPAPDPVLYOVNFNISDFNGKWY		1167
Db	2167	-KCVDIYKDNSSSTDGVETL-----	KQFYSKILTFSSASIVONADTYSVNFAPH			2216
QY	1168	ITSGINTEPDAQDCOLHEHTEBEDKNLVGNISMRKIKTLDSCGFTTSAVQKRVQDPNPGV				1227
Db	2217	EKESIMAIIRD--	IKKELTIFHONSIDISIV-----	EGVQVNL-----	A	2252
QY	1228	LYNHNDEIHAQDDVYILSSKIKNEBEDY-----	IFVYGRGDA-----			1268
Db	2253	LYDKLNEKREMDLYANNISEYTKLOMEHSTDFPKMIEFLHGMETNNKSULEKEKKL				2312
QY	1268	-----	WDGAGVAVYTRSSV--	LPN--	SIIPLEKAKSIGRD--	1302
Db	2313	SVNDHMSMEKEMKKNOKLKTTPESYONINNISYVIAEAKVKTLEETDRDYGONQVIEEHK				2372
QY	1302	-----	FSTFIETDNPCGPEDALVERIEKTEVEGE--	RIYKEVEEILEEVEKEVEKV--	GR	1353
Db	2373	KOFISILIDRTN-----	ALMDIEIEFKKENNYMLMEVNFETIHRANDYIEITKNLYOAK			2426
QY	1354	TE-----	MTLFOR--	LAEFNLKODEENFARLKSKEEM-----	EPLDE	1390
Db	2427	TEYEOILENIKQNDMLQNTFLKRSVIEYFENVKRRKKSILINDYIEDORLIKIGEHLDE				2486
QY	1391	IKMEASE	1397			
Db	2487	IKRNVTE	2493			

RESULT 5

hypothetical protein ORF64 - saimirine herpesvirus 1 (strain 11)

A; Note: host *Saimiri sciureus* (common squirrel monkey)

```
C:\Accession: H36812
C:\Date: 18-Oct-1992 #sequence_revision 10 Oct 1992 #ccac_change 00 00p 2002
C:\Accession: H36812
```

R;Albrecht, J.
submitted to the EMRI, Data Library, January 1992

A; Description: Primary structure of the herpesvirus salm1r1 genome.

A:Accession: H36812

A;Residues: 1-2469 <Al.5>

R; Albrecht, J.C.; Nichol

A;Title: Primary structure

A;reference number: A3/3
A;Contents: annotation;

A;Note: neither protein
C:Genetics:

A;Gene: 64

.....

1

1

1

1

1

QY	271	FTRSAVOTFVODBDLFGALYNHNEFLAHVODDWMYLLSSQTEPNKPDYIFVYVY-----	324
Db	178	FSTRSNVDAIEYSPSPNOYVLT--GSLYFVPRKYIGHSH-----YIMNHVIVYEXKL	228
QY	324	-GNDAMDGVGGSVYI-----TRSPILPESIIPLNLOKAASVGRDENFI	367
Db	229	HGHNIDLTQOEGILLISPPNTPKPTSTOKPXPPTPRKPAIPKAKTRPKRPTKESTI	288
QY	368	TJDNSCGPEPPLVERLEKTAEEGEKLLIEAV-----FIEEVEKEVAVRT--EMTL	419
Db	289	PYDKS--KKRP--KIPKTSKSKKVTIMDTALTPQHKHIEELHRELLPPIETRYVDNLT	343
QY	420	FOR-----LLEGRELO-----QDEENFVRELSKEKETLNFLOMEATEVERKL	462
Db	344	FHHVERTTGGIDSLTSLGINSTTKREDDEDDDNVTSKREDEDDWIDP-----	393
QY	463	FGRALPFRKLMAALPASHNPLANHEIKYVGSKLPGHFRSGWMDYGSIVYA-----	518
Db	393	---IPREVLDIETHS---DQELI-YMIGE-NHDMWSYDDDDIDIDISTQID	441
QY	518	-KICSSRRIPRFRKSPRCCGLDSGLDLSHGK-----HNLSPASHINQNPVG	567
Db	442	NITISLDNIPKN--NTEPR--IDKTSNPIKEGKALSHIDRLKNIVLEHILTISSIS	497
QY	568	NSGCKFPKVALAMVEKWOOFAKTATVAFIILSVASKAVALAKTCOTLLECRLEIAK	627
Db	498	ISKSKSLQFVL---WE-----KSIPIR-DLKITIKELITETIAELATK	542
QY	628	CISNPACAAVACLOTGNRPDETCQIKGDLFENSVADEFNECAVSRRKCVPRKSDVG	687
Db	543	-LTN-----DIFRNVNITLKNKCMALK-----SESDV	569
QY	688	DEP-----VPPDSVAVOKEDM-----KDFS-----GWFITRGINPTE	720
Db	570	SVKHLISALLNNILIKIQTIDIELKLTSTVETSELGKDFSVYCTKKESETIMAIKNIK	629
QY	721	DAFQOCLHEFHEENKVLGNLSWRJRTPDGCFTRTSAVOKF-----VODPKYVG	769
Db	630	EKISTRKQELHEENYFOSVL-----IAMEFQPIPLPTVRIQPSKKAQ	675
QY	770	ILYHNDNEVILVODMWYILSKVENSPEDIYFVYKGRND--ANDGYGGSVLYRSAYL	826
Db	676	QL--HEKSLVBOK---LTIDANVLDLHTMOKDIDISPAD-----FTTVLYKN	722
QY	827	PESIIPELQTAOKVGRD--FNFIKTDNTCGPEPPLVERLEKKEVEGERTI--IKEEVE	882
Db	723	IQSTLQLOTCVTIDINIKKFIISNVQOISYIGWVAELSHQMNFPADPVIPLKIID	782
QY	883	IEEVEKVRDEVTLFSKLFEFGKELJORDDEENFLELSKEEDVDYLGKMETEVEKLEFG	942
Db	783	IKKELQVTTTQ---KNEFTLSKILADVOTLLNMAQSDILSLPILOHYITKAGTLVG	837
QY	943	RALPIRKLMAVATHCFTSPCHDRIREFSSDDGIGRLGIT---RKRINGTELLKLTPEIQ	998
Db	838	E-----RENQFESLKNVQSLSEEFKLITLIDSTLENVQIDQ	878
QY	999	S-ADLRTGGRSSRPLSAFRSGFSKG--IIDIVLPSKNELKELTAPLL-----	1046
Db	879	EISDLQSNQVHIHSETIKQAFPRDKSNITIIINIIOLINQOKRTTYTQZMLLAVRFLSEA	938
QY	1046	-----KLIVGLACAFLLIVSADAVADATKCACLLKGRIELAKCIANPACANAVC	1096
Db	939	KFRESENTICETISILVSGSLTSKSTVEALKDALKSIDLTLEKUL-----	984
QY	1097	LOTENNRDETE-----CQIKGDDLFPENSVADEFNCAVSRRKKCVPRKSDLGEFPAPD	1149
Db	984	--TAVDRKRELINVIRKLOKLTLLLEQGFEDW-----KMEVJSF--VPT	1027

RESULT

6
A37102 myosin beta heavy chain, cardiac and skeletal muscle - human
N:Contains: myosin ATPase (BC 3.6.1.32)
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 20-Mar-1998
C:Accession: A37102; S12733; A94224; B28908; A24997; A27658; I54254; S12458; S09331;
R:Jelenke, T., Diederich, K.W.; Haas, W.; Schleier, J.; Lichter, P.; Pfordt, M.; Baer
Genomes 8, 194-206, 1990
A>Title: The complete sequence of the human beta-myosin heavy chain gene and a compar
A:Reference number: A37102; MUID:91065634
A:Accession: A37102 ;
A:Molecule type: DNA
A:Residues: 1-1935 <JAE>
A:Cross-references: GB:M57965; GB:M30603; NID:g179507; PID:g179508; GB:M30604; GB:M30
R:Liew, C.C.; Sole, M.J.; Yamauchi-Takahara, K.; Kellam, B.; Anderson, D.H.; Lin, L.;
Nucleic Acids Res. 18, 3647-3651, 1990
A>Title: Complete sequence and organization of the human cardiac beta-myosin heavy ch
A:Reference number: S12733; MUID:90301496
A:Accession: S12733
A:Molecule type: DNA
A:Residues: 1-106; 'E', 108-671, 'LYH', 675-857, 'A', 859-941, 'NV', 944-1123, 'A', 1125-1158, '
A:Cross-references: EMBL:X52889; NID:g29726; PID:g29727
R:Yamauchi-Takahara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
Proc. Natl. Acad. Sci. U.S.A. 86, 3504-3508, 1989
A>Title: Characterization of human cardiac myosin heavy chain genes.
A:Reference number: A94224; MUID:89264452
A:Accession: A94224
A:Molecule type: DNA
A:Residues: 1-87, 'Q', 89-106, 'E', 108-177, 1325-1702, 'DR', 1705-1786, 1788-1803, 'E', 1804-1
R:Yamauchi-Takahara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
Proc. Natl. Acad. Sci. U.S.A. 86, 7416-7417, 1989
A:Contents: annotation; erratum
R:Kurabayashi, M.; Tsuchimochi, H.; Komuro, I.; Takaku, F.; Yazaki, Y.
J. Clin. Invest. 82, 524-531, 1988
A>Title: Molecular cloning and characterization of human cardiac alpha- and beta-form
human atrium.
A:Reference number: A92770; MUID:88299163
A:Accession: B28908
A:Molecule type: mRNA
A:Residues: 1412-1518, 'R', 1520-1574, 'NV', 1577-1935 <KUR>
A:Cross-references: GB:M21665
A>Note: the authors translated the codon AGC for residue 108 as Arg
R:Lichter, P.; Umeda, P.K.; Levin, J.E.; Vosberg, H.P.
Eur. J. Biochem. 160, 419-426, 1986
A>Title: Partial characterization of the human beta-myosin heavy-chain gene which is

A:Reference number: A24997; MUID:87030293
 A:Accession: A24997
 A:Molecule type: DNA
 A:Residues: 682-721;975-1112;1854-1935 <LIC>
 A:Cross-references: GB:X04627
 R:Saiz, L.J.; Gianola, K.M.; McNally, E.M.; Peghali, R.; Eddy, R.; Shows, T.B.; Leinwand
 Nucleic Acids Res. 15, 5443-5459, 1987
 A:Title: Human cardiac myosin heavy chain genes and their linkage in the genome.
 A:Accession: A27858
 A:Reference number: A93669; MUID:87260010
 A:Molecule type: DNA
 A:Residues: 1854-1865, 'A', 1867-1935 <SAE>
 A:Cross-references: GB:X05631; GB:X00362; NID:934643; PID:934644
 R:Diederich, K.W.; Eisele, T.; Ried, T.; Jaenicke, T.; Lichter, P.; Vosberg, H.P.
 Hum. Genet. 81, 214-220, 1989
 A:Title: Isolation and characterization of the complete human beta-myosin heavy chain ge
 A:Reference number: 154254; MUID:89154425
 A:Accession: 154254
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 653-720 <RES>
 A:Cross-references: GB:M27636; NID:9179511; PID:9601916
 R:Bober, E.
 Submitted to the EMBL Data Library, January 1989
 A:Reference number: S12458
 A:Accession: S12458
 A:Molecule type: mRNA
 A:Residues: 785-1076, 'E', 1078-1123, 'A', 1125-1702, 'DE', 1705-1935 <BOB>
 A:Cross-references: EMBL:X51591; NID:929467; PID:929468
 R:Bober, E.; Buchberger-Seidel, A.; Braun, T.; Singh, S.; Goedde, H.W.; Arnold, H.H.
 Eur. J. Biochem. 189, 55-65, 1990
 A:Title: Identification of three developmentally controlled isoforms of human myosin hea
 A:Reference number: S09331; MUID:90235862
 A:Accession: S09331
 A:Molecule type: mRNA
 A:Residues: 785-830, 'X', 832-900, 'X', 902-970, 'X', 972-1040, 'X', 1042-1076, 'E', 1078-1110, 'X',
 '1602-1670, 'X', 1672-1702, 'DE', 1705-1740, 'X', 1742-1810, 'X', 1812-1935 <BOB>
 A:Cross-references: EMBL:X51591
 R:Jandreski, M.A.; Liaw, C.C.
 Hum. Genet. 76, 47-53, 1987
 A:Title: Construction of a human ventricular cDNA library and characterization of a beta
 A:Reference number: S02229; MUID:87192738
 A:Accession: S02229
 A:Molecule type: mRNA
 A:Residues: 1393-1702, 'DR', 1705-1935 <JAN>
 A:Cross-references: EMBL:X06976; NID:934860; PID:9825694
 R:Saiz, L.; Leinwand, L.A.
 Nucleic Acids Res. 14, 2951-2969, 1986
 A:Title: Characterization of diverse forms of myosin heavy chain expressed in adult huma
 A:Reference number: A93616; MUID:86176778
 A:Accession: B23767
 A:Molecule type: mRNA
 A:Residues: LUGVGLASG, '1311-1312, 'G', '1314-1355, 'R', '1357-1358, 'GD', '1361-1438, 'LO', '1441-
 A:Note: the first ten codons of the sequence figure show the reverse complementary stran
 A:Genetics:
 A:Gene: GDB:MYH7
 A:Cross-references: GDB:120215; OMIM:160760
 A:Map position: 14q12-14q12
 A:Intons: 67/3, 115/3, 168/1, 177/2, 213/3, 244/3, 266/1, 299/1, 333/3, 380/1, 419/3, 4
 24/3, 1390/2, 1451/3, 1507/1, 1548/3, 1651/3, 1719/3, 1761/3, 1853/3, 1930/3
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methylate
 F:88-766/Domain: myosin motor domain homology <MMOT>
 F:178-185/Region: nucleotide-binding motif A (P-loop)
 F:348-585/Region: actin binding #status predicted
 F:655-677/Region: actin binding #status predicted
 F:839-1935/Domain: coiled coil #status predicted
 F:839-1279/Region: S2
 F:1280-1935/Region: light meromyosin
 F:129/Modified site: N6, N6, N6-trimethyllysine (Lys) #status predicted
 F:184/binding site: ATP (Lys) #status predicted
 F:695,705/Active site: Cys #status predicted

Query Match 1.9%; Score 146; DB 1; Length 1935;
 Best Local Similarity 16.9%; Pred. No. 0.74;
 Matches 238; Conservative 207; Mismatches 529; Indels 434; Gaps 51;

OY	203	DVGEFPPVPRNAVVO-----	-----NFMKDFSGKWITSGLNTPDPAFCQHEF	248		
DB	461	DIAGFEIFDFNSFEOICINFTEKLOQFFNNHMFLEQEEYKKEIEWTFIDFGMDLOAC	520			
OY	249	HMDNDLVG-----	-----NLTKRIKLDGSGFTRSAGVQVFOVDDPDLPGALYHND	293		
DB	521	IDLEIKPMQIMTLDEECMPKATMTFPAKLFD-----	-----NHL	557		
OY	294	NEFLHTQDDWYILSSQIEKKPD-DYLFVYRGRND-----	AMDGYSVYTRSPPLPESI	348		
DB	558	GKANOK-----	PRIRKGRPEAHSILHYADIVYNIIGW-----	LQKKDKDLNTEV	605	
OY	349	IPNLQRAA-KSYGRDENFTTDSNGCEPPPLVERLEKTAEGEKLILKEAVEIEVEK	407			
DB	606	VGLYQKSKLSTLTFANY-----	AGADAP-IEKGKAKKSGSFQVVSALH-RENLNK	657		
OY	408	EVEKVDTEKTLFQRLLEGFKELQDEENFVELSKEE-----	ILNELQMEAT-	458		
DB	658	LMTNRST-----	HPHFVGCILIPNETKSPGVNDPLVMHQNCNGL	699		
OY	458	EVEKLFGRALP-----	IRKLRLALPHSNFLANHETIKYVGSKLPGHKRFSWG	506		
DB	700	EGIRICKRGPNNIIXGDFRQRRILNPAALBGOIDSRKAEEKLSLDIDHNOYKFG	759			
OY	507	WEDYFGSIYVATCGSSRRIPRFRKSPRICGIDSGLOLFSGKHNLSPANS---	INON	563		
DB	760	HTVVFAGALGLLEMPRELSRITRI--	QAOSRGVLAEMKYKLLERRSDSLVITOWN	817		
OY	564	VPRGNSCKRFPKVALMWMKMGQFAKTAIVAFILSVASKDAVDALTKCTCLKEC--	622			
DB	818	IR-----	ATMGVKNM--PMKLIFFIKPLKLSEREKMAKMEETRLKEALE	864		
OY	622	-----RLELAKCISNPACANVACLOTNNRPDETCQIKCGDLFENSV----	VDEFNE	671		
DB	865	KSEARRKELEKKNVSLLOEKNDLOVQAEDNLADAEERCDLKNKIQLEAKYEMNE	924			
OY	672	-----CAVSFK--	KVPRKSDVGPDPVDPVSLVQKFMKPFSGKWFITRBL	716		
DB	925	RLEDEEMNAELTAKRKRLDECESELRDIDDELTLAKYERK-----	969			
OY	717	NPTFADACQDHEPHFEENKLVGNLSWRIRTPDGFRTSAYOKFVODPKYPIILYHND	776			
DB	969	-----HATENK-VKNLIEEMAGLD-----	ELIKLTKREKALQEAHQAL	1007		
OY	777	EYLLYDDWYILSSKYNSPEYIFVYKGRNDAMDGYSGLYTRSA-----	825			
DB	1008	DDLQAEEDKQNTLTKAKVLEQOV-----	DDLESLRQEKKVRMDLERAKKLE	1056		
OY	825	---VLPSSIIPELQTAOKYG-----	RDF-----	NFIKIDNCGEB--	PRIVE	863
DB	1057	GDLKTQESIMDLENDKQODERLKKRDFELNLMARIEDQALGSLQLOKIKLEIQARLE	1116			
OY	864	RLEKVEEGERT-----	IKREVEIEEVEK-----	890		
DB	1117	ELEBEL-ESEETAARAKAKELRSDLRLEISRLDEAGATSVQLEMMKRAAEFOKMR	1175			
OY	890	-----VRDEVTLLFSKLFGEKELQDEENFRELSEKEMDVLDGKME	933			
DB	1176	RDLBEATLQHEATAAALRKKNHADSVALBQINDLQVKKLKEKSEFEKLIDD-----	V	1231		
OY	934	ATEVEKLFGRALPFRKLMANATHFTSPCDRIREFSSDGIORLGTTRKRIINGTFLAKI	993			
DB	1232	TSNMEQIIKAKANLEKM-----	CRT--LEDQMEHRS-----	KAETQSVNDL-----	1274	
OY	994	LPPIQSADLRTGGRSSRPL--	SAPRSGFSKGIFFIVPLPSKNEKELTAAPLLKLVGV	1050		
DB	1274	--TSQRAKLTQENFELSRLQDEKFAALLSQLTRG-----	KLTLYQLEDDKRLQEEVEYKAK	1326		

QY 1051 LACAFLEVPASADAVDALKTACLLKGGRIELAKCIANPACAAVACLOTNNRPDETEQ 1110
 Db 1327 NALAHALQSRHCDLIREYEETPAKAELOVLSK--ANSEVAQWRT----- 1374
 QY 1111 IKCGDLFENVSEVDFNECAVSRKCVPRKSDLGEPAPPSVLVQFNISDFNGKWTYS 1170
 Db 1374 ----KYEETDAIQRTTELEHAKKKLAQRLQEAEE-----ANE 1405
 QY 1171 GUNPTDAFPCQALHEFTEDDNKLVGNISKRITLDSGFTTSANQKEVQDPNPGVLYN 1230
 Db 1406 ANAKSSSEKTKHRLQNELTEDLMVD-----VERSNAAMALDKKQ-----RN 1448
 QY 1231 HDNEVLHYODDWYLLSSKLENKPEDEYFYVYRGNDAMDYGGAAYVTSSVLPSIIEP 1290
 Db 1449 FDKILAEWQKYEESOLESSQKE-----ARSLTEFLKNAVEBS 1491
 QY 1291 LE-----KAASIGRDFSTFRTONTGCPPEALVERIEKTYEEGRILVKEVEIEEY 1344
 Db 1492 LEHLETFEKKENKMLQELISDLTEQLSGSKTTHELEKVKQLAEKMELOSLAEKAST 1551
 QY 1345 EKEVEKVGTEMTLFO-----RLAEGFNLKODEENVRELS----- 1382
 Db 1552 EHHEGKILRAQLEFNQIKAEIEKKAKEDEWEQAKRNHLRVAVDSLQTSLSAETRSEEA 1611
 QY 1382 ---KEEME-FLDEIKMAEVEKLEFGKA 1405
 Db 1612 LRYKKKEGDLNEMETOLSHANRMAEA 1639
 RESULT 7
 S57976
 nucleic acid migration protein NDM1 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YD8358.06; protein YDR150W
 C:Species: Saccharomyces cerevisiae
 C:Date: 13-Jan-1996 #sequence revision 01-Mar-1996 #text-change 06-Feb-1998
 C:Accession: S57976; S19052; S17018
 R:Murphy, L.; Richards, C.; Harris, D.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: S57976
 A:Accession: S57976
 A:Molecule type: DNA
 A:Residues: 1-2748 <MUR>
 A:Cross-references: EMBL:Z50046; NID:G899393; PID:G899399; MIPS:YDR150W
 A:Experimental source: strain AB972
 R:Kornatowicz, J.; Schaaf-Gerstenschlaeger, I.; Zimmermann, F.K.; Percebo, D.; Kuentzel, H.
 Mol. Gen. Genet. 230, 277-287, 1991
 A:Title: Nuclear migration in Saccharomyces cerevisiae is controlled by the highly repet
 A:Reference number: S19052; MUID:92079907
 A:Accession: S19052
 A:Molecule type: DNA
 A:Residues: 1-1569, 'V', 1571-1821, 'K', 1823-1959, 'RHL', 1963-1970, 'RN', 1973-2048, 'N', 2050-2
 A:Cross-references: EMBL:X61236; NID:94071; PID:94072
 C:Genetics:
 A:Gene: SGD:NDM1
 A:Cross-references: SGD:S000255; MIPS:YDR150W
 A:Map position: 4R
 Query Match 1.9%; Score 146; DB 2; Length 2748;
 Best Local Similarity 17.9%; Pred. No. 1.2;
 Matches 283; Conservative 232; Mismatches 581; Indels 486; Gaps 72;
 QY 62 SFSFSSCHCKRQSCIDTSFEELQFRLKRGMTLLEKOWROF-----IQLAIVLC 115
 Db 748 SAYEDLVCKENPDV-----EFLKEKSAKLGHIVVSEYSELQKRYSELEKEVQPS 800
 QY 116 TFEVIRPVDADALKTACLLKEGIEELAKCIANPSC-----AANVACIOTCNRPDDET 169
 Db 801 LAYIVHAHAKTDH-----HLSDSAVEELVCKENPDMEFLKEKSAKLGHIVVSNFAYSEL 856
 QY 170 ECGI-----KCGD--LFENSVVDQFNCAVSRKCVPRKSDLGEPAPPSVLVQFNISDFNGKWTYS 215
 Db 857 EKKLEQPSLAVIVHAHAKTDHHLSDSAVEELVCK-----KENSIV-EE----- 900

QY 216 VQNMKDFSGKWTITGLNPTDAFDCQALHEFTEDDNKLVGNISKRITLDSGFTTSANQKEVQDPNPGVLYN 275
 Db 900 ----LKEKSAKLGHIVVSNFAYSELKLEQ-----PSLAVIHAHAKTDHHLSDSA 948
 QY 276 VQTMQPDLPALYNHNEFLHYQ-----DWTYISSQIENKPPDDYIFYRNG 324
 Db 949 YE-----DLVCKENPDMEFLKEKSAKLGHIVVSNFAYSELKLEQPSLEYIVHAHAK 1002
 QY 325 RND---AMDYGGSV-----LYTRSPULPELIPN-----IQKAKSGRPF 364
 Db 1003 TNNHLLSDSAVEELVCKENPDMEFLKEKSAKLGHIVVSNFAYSELKLEQPSLEYIV 1062
 QY 364 ----NNEFTTDS-----CGPEPLVERLEKTAEGEKLLIKEAV-EIEEVEKE-- 409
 Db 1063 HAKATNHLSDSAVEELVCKENPDMEFLKEKSAKLGHIVVSNFAYSELKLEQPSLEY 1122
 QY 409 --VEKVDPTMTLPQRLBEQFLOQDEBNFVRLSKEREITLNE--LOMEA-TEVEKLE 463
 Db 1123 YIVHAHAKTDHHLSD--SAVEELVCKENPDMEFLKEKSAKLGHIVVSNFAYSELK 1179
 QY 464 GRALPTIKLMLALPSPNFANHETIKYVGSKLPGRKPFSGMEDYFGSIYVAR----- 519
 Db 1179 -----KLEQ--PSLAVIHAHAKTDHHLSD--DSAYEDLVCKENPDV 1217
 QY 519 -----ICSSRIPIRYEKSPPICGLDSRGL-OLFSHKNLSPASHINOV 564
 Db 1218 EFLKEKSAKLGHIVVSNFAYSELKLEQ-----LEQPSLAVIHAHAK--ATDHLSDSA 1268
 QY 565 PKNSGCKPEPKDVALMWKMGQFAKTAIVAFILSVAKAD-----AVDALTKICL 617
 Db 1269 YEDLVCKENPDMEFLK-EKSAKLGHIVVSNFAYSELKLEQPSLEYIVHAHAKTDHHL 1327
 QY 618 LKRCLE-LAKCISNPAC-----AANVACIOTCNRPDDETEQIKCGDLFENVSEVDFNE 670
 Db 1328 LSDSAVEELVCKENPDMEFLKEKSAKLGHIVVSNFAYSELKLEQPSLEYIVHAHAKTDHHL 1378
 QY 671 ECAYSRKKCVPRK-SVYGF-----PVDPVSVLVQKEDMKDFSGKWPITRGLNPTDAF 723
 Db 1378 EYLKHAEOIQSKTISISDFNTLANPSMEDMAKSLQKLEYQIVSDEYI--ALKNTMEKP 1435
 QY 724 DCQ-----LHEFH-----TEENKLVGNLSWIRITPDGFFTRSAVQKFOVD--PKYGIILYN 773
 Db 1436 DVELLSKLGKHIIDTITYNELVSFN-----SPTLKEFLKEKSAKGYRLI 1482
 QY 774 HDNEYI-----LYQDDWYILSK-----VENSPE-----D 798
 Db 1483 EPNEYIDLNRITATPESKEIDNFCQIGYALDSKEYERLKNLSLENPSKFFIEENALLD 1542
 QY 799 YIVP---YKGRNDAMDYGGSVLITRSVILPELSIIPELQTAQKVGDPNFTFTDNTC 855
 Db 1543 LVYVDTEYQAMKD-----NASKNSKSLIPSTKAL-----DFVTM----- 1577
 QY 856 GPEPLVERLEKVEGERTIIEVEEIEEVEKVRDEVLFLSKLFGFGELOD----- 912
 Db 1577 -PAPQLASAKESLSQK-----RILSDIENELKAL-----GYAIREKNENPN 1616
 QY 912 -EENFIRELSKEDM-----DVLGKMEATEVEKLEFG-----RALPIR 948
 Db 1617 LKPIVDNASKNDVNLCSKPSLVPSTEEDYDNMRKEHTKILNIGDPSIDFLKEKCEKY 1676
 QY 949 KLMVAATN-----CFSPCHDRIRFSSDGIQRLQITKRKINGTFLAKILPPIQSD 1001
 Db 1677 QMLIISKHYEQKQAIENPGYEFLEKASALGY-----ELVSEVLDKMKQMIDSPD 1729
 QY 1002 LRTTGRSSRPISAFRSQSGKIPDIVLPKSNELKELTABLLKIVGLVLAACALVISA 1061
 Db 1730 IDYMQEKAARN-----EMVLL--RNEKE--ALQKRIEYPLTFLIEKAA 1770
 QY 1062 -----DAVDLKTACLLKGRITELAKCIANPACAAVACLOTNNRPDETEQIKCG 1114
 Db 1771 GNNKILVQIEEDEN-----IRKC-----NHPTRMELEESCH 1802

Matches 268; Conservative 275; Mismatches 556; Indels 500; Gaps 71.

```

QY 44 RSNNGVFNFRFTSYKTSFS--DSSHCKKQKQISIDTSEELQRPDKGM--TLLI 98
Db 482 RFNNNTSNLNDINAYTALPESLSSSSKFEOMESKYSFTD---KLAGMEFSLM 537
QY 99 LEKORFOIOLALVCTEVIIVPRVADALKTACCLKECRLELAKCIANSCAANYAC 158
Db 538 YGEKFEFLSQ-----EATNNYQEPDLNKKLNE----- 567
QY 159 LQTCNNRPDETCQICGGLFENSVDQFNCAVSKKCVPRKSDVGEFPVDRNAVYQN 218
Db 567 IESFYMMFKTOETLK-----VD-FNTSLIN-----IKDEIGK----- 599
QY 219 FNNKDFSKWYITSGNPTFADQCLHEFHENDKLVGNLWTKITLDGFTTSAYOT 278
Db 599 -NVEERORY-----DEVNIFVTOLESKLQYKSKWQEMDSNKLKE----- 641
QY 279 FVQDDPLPGALYNHNDNEFLH---YDDWYILSQIENKPDYIFVYRRGRNDADWGYG 334
Db 641 -----SQINKTNEFSLIQIOQDKGIELSESVPNDLSHI-----QKKALDHGS 686
QY 335 ---SVIYTSPL-----PESIIPIQAKSVGRDPNNFTTDSGCEPPYERLE-- 385
Db 687 WKDELIALKSLDLIKVSSEELLSATLIKIESLEKDVN-----DRMEYV 730
QY 385 --KIAEESEKLLIKEAVE-----IEEVEKEVEKYRDTENTMLFORLEGF 427
Db 731 LKTGTGI-ESVIEKKEKELKMSYSDPALLIGEFIRQIETIKKDSVPFLELDNKF 789
QY 428 KELQDEENFV-----RELSEKEELINELQMEATE-VEKLFGRALPIKILMA 475
Db 790 ---DDKNFVYSKIEEDCYKDKFKIESEDILNNFKSDPLNEFIRS-----KLQIV 836
QY 476 LAPHS-----NFLANHEITKYVYSGSKLGHKRFSGMWDYGSIVY--AK 518
Db 837 SNIKSDNKOIODEFDRIKSDILNKRKDSINNEVDSKLS-----DWQSKLNEITVKIEN 889
QY 519 ICSSRRIP-----RYFRKSPRICGLDSRGLQFSHGK----- 552
Db 890 LSSGKVDLDLIDSEYTKIKELKFSIS-----LESYILEKIDEFNOCAGIYSDLEQ 943
QY 552 ---HLSAHSINONVRG-----NSGCKPKVDALVWEXWGOFAK-TAIVAFILISV 601
Db 944 DIMNHFNETRELEENLSKFAAVLNSEEFYKEDSLQDKRIDIASFOANIDITDLSL 1003
QY 602 ASKADAVALKTCTCLKECRLELAKCISN-PACAAVAVAC----- 641
Db 1004 NKRENDIN-----KEINKYNEVISNRYGSENISSKLENIHAEINLSRRLTDR 1054
QY 641 -----LQTCNNRPDETCQI-----KCGDLFE-----NSVDFEFCVAVSR-K 677
Db 1055 IDSLSKGDENLOKLESFVASKYQVEKELKVLDLTDGGEAKINKLKVLEIQQYKKSLE 1114
QY 678 KCVPRKSVGDFPVDPVSVLVQKFMKDFSGKWFTRRLNLTTPDAFCQLHEFTEENKL 737
Db 1115 EADIRRTI-----DNDIM-----QAKERFGE--ITNELKNNISSEKSEFLNDLYKEREKFL 1162
QY 738 V-GNLSMIRIPTDGGFTFRS-AVOKFVQDDPKYPIILYNHNEVELLYQDDWYILSSVEN 794
Db 1163 IESNEEERYST-----FLIESEGAISK-IRDEITYKTLTISNDEN-----LQIKISE 1206
QY 795 SPEDYIENYVYGRN-----DAMDYGGSVLYTRSAYLPESIIPELOTAQKV 841
Db 1207 MDONEEIIEOGRSKDLEFEKELEODIKDCY-GFINSGEIKAGVEENIKNHFPVCIKV 1265
QY 842 GRDFTETIKLNTGCEPPYERLE--KYPEGERTIIKEY-ELEIEVEVVRD---KE 894
Db 1266 ---NTLI-DDDIKYENELIKRIDSLSKSIESTDSEIEKNLNDVSGCIDKIDANDPMUKY 1320
QY 895 VTLSEKLEGEKRELQDRENFLEISKEEMDVLDGLKMEATEVEKLFGRALPIKILMAVA 954
Db 1321 IELERECNEGQLENLKNKIDNKIKALDINALSQYDGLEKKYADKIDFESERL--NSYIATL 1378

```

```

QY 955 THCTSPCHDRIRFFSSDDIGRLGITRRKING-----TFLKLIPLPIO 998
Db 1379 SEEFKSNKEMI--FELESQKMLKLNLESDLNVEKDVIRLKEESYHANSYHKLLEEDF 1436
QY 999 SADRITGGRSSRPLASFRSGFGSKGIFDIYPLPSKNEKELIAPLLKLVYLACAFILIV 1058
Db 1437 FKDKINGELKSLLENFIASTYNDKIQNLLEYLDSKN----- 1473
QY 1059 PSADAVADALKTACCLKGRLELAKCIANPACAAVAVACLOTCNNRPDETCQICGDLFE 1118
Db 1473 -----LENKTELLQSPRLDIEQKMDKDENFYIDFTKEESKSKKMDQSEIA---LME 1521
QY 1119 NSV---VDFENCAVSKKCVPR-----KSDGEPAPDPVSVLVONFNISDF-----NG 1164
Db 1522 TNTGKRVDEFVDPVNNKOSLIDSMFLNIDVDYKDMQEKSYSTIERKINLAFGIKSFEND 1581
QY 1165 KWTITSGLNPTPAPFCOLHEFTEGDN-----KLWGISWRIKTLDSGFT 1211
Db 1582 IFNVKIGLESFKGFEIKAEIIFSNLQNAKKIEOSVHLDFRNIG-SUNLKVLDENFV 1640
QY 1212 RSVQKFEVQDPNQ-BGVLYNHDEYLAHYOD-----WYILSKTIENRPE 1254
Db 1641 DFKLEKIDKVNKKTEDILIQAEVKFLQOKDLEKIFELNOKLHEFTTSSNNDKYRR 1700
QY 1255 DYIFVYRRGNDAWDYGGAVYTRSSV-----LPNSIIPLEKAAKSIGRD 1301
Db 1701 EMYDVLSSDK---ESFEQIELINKNISFSEKISLYNNIETISENYSFSKISKD 1756
QY 1302 FSTFIRTDNCGEPALVERIEKTEVEGERIIVKVEEIEEVEKEVEKVGRTENTLFO- 1361
Db 1757 LG-----LLEDELKKSUKHS---TSEETIKSLQDQIDK-----FEV 1791
QY 1361 RLAEGFNEKQDEENFVRLSKSEM-----EPLDEIK 1392
Db 1792 EFKKNHKEKLEKYDNNILESKILNCDOVQFNKFTSEIK 1830

```

RESULT 10

RESN-H3 antigen PFB0915w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 13-Nov-1998
 C:Accession: B71603
 R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600
 A:Accession: B71603
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1558 <GAR>
 A:Cross-references: GB:AB001424; GB:AB001362; NID:G3845307; PID:G3845309; TIGR:PFB091
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0915w

Query Match 1.9%; Score 143.5; DB 2; Length 1558;
 Best local similarity 17.9%; Pred. No. 0.77;
 Matches 206; Conservative 153; Mismatches 377; Indels 413; Gaps 50;

```

QY 347 STIPNOK-AASVGRDN-----NITTDNSGCEPPYERLE-----RLKTAPEG 390
Db 626 TIDNEVEETTAESVTFSSNILEIOENTITNDT---EKKLEELHENYLSALENTQSEE 682
QY 391 EKLIIKEAVEIEEVEKEV-----EKVDFTEKTLFORLEGEFKELQD---EENFREL 441
Db 683 EK---KEVIDIVEEVKEEVAITLIEVQAEBSASTITTELFENLEENAVASNNENAVNL 739
QY 442 SKDEEELINELQMEATEVEKLFGRALPIKILMALAPHSNFLANHEITKYVYSGSKLPGRH 501

```

```

Db 740 EKLNETVNTVDKVEEVEISGESLENNENDKAF-----FSEIDNVKGIQENLITG-- 793
QY 502 RESMGMEDVFGSIYVAKICSSRRIPRYPRKSPRICGLDSRGLQJFSHG---KHNLSPA 557
Db 793 -----MFRSIEFSIYQSE--KYDLENVVSIIIDN--IENKRGLLINKLENISST 840
QY 558 HSINONVPRG-NSGCKFPKVDALMYEKMGCPAKTAIYAFILSLVASKADAVALDKTCTC 616
Db 841 EGVQETVEIHEVQNYVDVDPAMK-----DOF-----LGLINEAGG----- 878
QY 617 LKECHCLLAKICSNACANACIOTCNNRPDETCQ---IKCGDLFNSVYDEFNECA 673
Db 878 -LKEMFENLDEVFKS---ESDVIYVEIETDEVOKEVEKEIYIIEEMENIVDVEE-- 932
QY 674 VSRKCKVPRKSDVG--FPVDPDSVLVOKFDMKDFSGKWFITRGLNPTFDACQLHFEH 731
Db 932 -----EKEDLTDMKIDAVESEIIS--DSKEET-----ESI 961
QY 732 TEENKLVNLSMRIRTPDGGFTRSVOKFVODPKYIPLIYHNDNYLLYODWYILSK 791
Db 962 KQKEDVSLVEEVQDND---MDESVEKYLE-----LKNMEELMKDAVEINDITSK 1010
QY 792 -VENSDEDIYFYKKRNDAMVGGSVLYTSAVLPEI--IPELOTAAQVGRDPTNF 848
Db 1011 LIETQE-----INEVADILIKMEIKELKFA---LSEDSKEI 1046
QY 849 IKTDNCGPEPLVERLEKKEVEGE--RTIIEKEVEI---EEVEKVD-----K 893
Db 1047 IDAKD-----DLEKVIIEEHDIITTLDEVVELKVOEDIKIKVSDLKMLEEDIK 1097
QY 894 EV-----TLSSKLEPEKRELODEENFLRELSKEMVLDGKMEALEVEKLGRAI--- 946
Db 1098 EVKEIRELESEIIEDEYKELTITDLEEKKELEKHFEFESEAEIKIDLEADILKEVS 1157
QY 946 -----PIRKMAVATHCFTSPCHDRIRFFSSDDIGLITRRKINGTFLKIL 994
Db 1158 SLEVEERKLEEYHLEKEVEHILSGDAH-----IKG----- 1190
QY 995 PPIQADLRITGGRSSRPLSAFSGSGKIFDIYPLPSKNELKELAPLLKLVGLA 1054
Db 1190 --LEEDLEEVDDLGKSTILMLKGDMELGDM-----KSLBDVIKTL----- 1231
QY 1055 FLIVPSADAVALKTACCLIKGCRILEAKCIANPACANACIOTCN--NRPDTECOIK 1112
Db 1231 -----GERVESLKDVLSALGMDDEQKTRKKAORPLLEVLK 1269
QY 1113 CGDLFNSVYDEFNECAVSRKCVPRKSDGEPAPDSVLVONFINISDFNGKWIYISGL 1172
Db 1270 -----EEVEEKPKKTKTKK-----VREDIKD-----K 1292
QY 1173 NPTFDACDOLHEFHTEGDKNLVGNISWRIKTLDSGFTTSRAVOKFVODPNQCVLYNH 1232
Db 1293 EPKQEIYEVKMKDDIDED-----IEEDVEED----- 1320
QY 1233 NEYLYHODWYILSKIEKNPEDIYFYRRGRNDAMVGGAVYITTSVLPNSIPELE 1292
Db 1320 -----IEEDKVEDIDEDID-----E 1334
QY 1293 KAAKSIGND-----FSTFRTDNTCGPEPALVERIEKTEVEGERIYKVEE-----IE 1341
Db 1335 DIDEIDIGDKREVIDLIVQKRIKIEKVEKKKLEKVEBSGLKHHVDEVMKYVQKID 1394
QY 1342 EEVEKEVEKV--GRTENTLIFORLAEGFENELKODEB-----NEYRELSKEEMELDEIKM 1393
Db 1395 KEVDREYKALSKENDVY-----NVLKONODEFSKVNFKYKXVFAAPISAVAA 1445
QY 1394 EASEVEKLE 1402
Db 1446 FASYVVGFF 1454

```

RESULT 11
S06006

```

myosin beta heavy chain, cardiac muscle - rat
N:Contains: myosin ATPase (EC 3.6.1.32)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence:Revision 31-Dec-1993 #text_change 13-Feb-1998
C:Accession: S06006; S07536; I67441
R:Krafft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
Nucleic Acids Res. 17, 7529-7530, 1989
A:Title: Complete nucleotide sequence of full length cDNA for rat beta cardiac myosin
A:Reference number: S06006; MUID:90016823
A:Accession: S06006
A>Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-1935 <KRA>
A:Cross-references: EMBL:X15939; NID:956556; PID:956657
R:McNally, E.M.; Krafft, R.; Bravo-Zehnder, M.; Taylor, D.A.; Leinwand, L.A.
J. Mol. Biol. 210, 665-671, 1989
A:Title: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Compari
A:Reference number: S07535; MUID:90133919
A:Accession: S07536
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-950, 'RR', 953-1935 <MCN>
R:Mahdavi, V.; Lompre, A.M.; Chambers, A.P.; Nadal-Ginard, B.
Eur. Heart J. 5, 181-191, 1984
A:Title: Cardiac myosin heavy chain isozymic transitions during development and under
A:Reference number: I5305; MUID:85179510
A:Accession: I67441
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1871-1935
A:Cross-references: GB:M32698; NID:9205598; PID:9205599
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methyl
F:88-766/Domain: myosin motor domain homology <MOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:548-585/Region: actin binding #status predicted
F:655-677/Region: actin binding #status predicted
F:839-1935/Domain: coiled coil #status predicted <COI>
F:839-1279/Region: S2
F:1280-1935/Region: light meromyosin
F:129/Modified site: N6/N6-Triethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:695,705/Active site: Cys #status predicted

Query Match 1.9% Score 142.5; DB 1; Length 1935;
Best local similarity 16.7% Pred. No. 1.2; Mismatches 558; Indels 459; Gaps 56;
Matches 247; Conservative 218;

QY 166 PDETCQIKCGRLFNSVVDQENECVSRKCVPRKSDVGEPVDRN-----AVV 216
Db 375 PDGTE-----EADKSAIIMGINSADLKGICHPKRVKGNEXYKGVQVQVAYAGLA 428
QY 217 QNFNMKDFSGKWIYISGLNPT-----PDAFDQLEH---FHMENDKL 255
Db 429 KSVYKEMFN--WMTNR--INATLETROPROYEIGVDIAGFEIIFENSFQELCINFTKEL 485
QY 256 VGNLWIRKITLNGSGFTTSRAVO--TFVQ-----D 282
Db 486 QOFFNHMHVFLVQOEYKKGIEWTFIDFGMDLQACIDILIEKPMGMSLIEBCMFPRATD 545
QY 283 PDLPGALY--NHUNEFVHYDDWYIILSOTENKPDYIFVYVYGRND-----AMDYGGGSI 337
Db 546 MTFKATLYDNHLSKSNPFKPNINIKQDAH-----FSLIHVAGTVDYNIIGW-----L 594
QY 338 YTRSPILPESILPNIQKAA--KSVGRDPNNFTTDSNCGPEPLVRLKTAEGEKLILK 396
Db 595 QKNKDPINETVYVLYOKSSLSKLISNLFANY-----AGADAP--VDKGGKAKKGGSSQTV 647
QY 397 EAVEIEEVEKEVEKVRDTEMILLFQRLLEGFKELQODEENFVRELKSEKE----- 448
Db 648 SALH--RENINIKITNLRST-----HPIVRCIIPNETKSPGYMNDPL 688

```

QY 448 ILNLOMEAT-EVEKLEGRALP-----IRKLMALAPHSNFIANHEITKYVVS 495
 Db 689 VMHOLRCMGVLEGIIRICRKGPNRIILYDGFROQRILNPAALPEGGFIDSRGAELILOS 748
 QY 496 KLPGRHRSWGMEDYESSIVAKICSSRIIPRFKRSPPICCGDLSRGJOLFSGHKNLS 555
 Db 749 LIDIDNOYKFECHTVKFKAGILGLEEMRDERLSRIITRI--QAQSRGVLSRMEFKILLE 806
 QY 556 PAHS---INQVVPKNSGCKFPKDVALLMVEKMGCFAKIAYAFILTSASADAV----- 609
 Db 807 RRDLLIIONNIR-----AFMGVKNPWRMKLYFKIKPLKLSAETEKEMANKE 854
 QY 609 -----DALTCCTCLKECRLELAKCISNPACANVACLOTQNNRPDETCOIKGDLFE 662
 Db 855 EFGVAKDALKESEARKELEEKVSLQJ-----KNLOQVOQAEDQNLADEERDOLIK 910
 QY 663 NSV-----VDEPNE-----CANSRK---KCPVRSKDVGDPPVDPBSVLVQKED 702
 Db 911 NKIOLEAKVEMTERLEDEEEMNAEILTAKRKLEDECSLKRDIIDLLELLAKVEKER-- 969
 QY 703 MKDFSGKWFITRGINPTFADFQCLHEPHEENKLVGNLSWIRTPPDGFTFRSAVQKRV 762
 Db 969 -----HATENK-VKNLTBEMAGLD-----ELIVELT 993
 QY 763 QDPKYGILYHNHNEYLLYQDDWYILSKVENSPEDEYIFVYKGRNDAMDGYGGSVLYTR 822
 Db 994 KKKKALQEAHQALDLOAEEDKVNLTAKAKVLEGOV-----DLEGSLODK 1042
 QY 823 SA-----VPESTIIPLOTAQKVG-----RDF-----NFIKTIDNTGCP 857
 Db 1043 KVMMDLERAKRLEGLDKLTQESIMDLNDKQOLDERLKKKDFEINALNARLEDEQALGS 1102
 QY 858 E-----PVLVERLEKKVEEGERT-----IKEVEELEEVEER----- 890
 Db 1103 QLOKKLKELOARHEELEBEL-EABRTAKAKVKLSDLSRELEETSELEERAGATSVQI 1161
 QY 890 -----VBDKEVTLFSKLFEGFKELQDEENFLREL 919
 Db 1162 ENMKRREAEFOKKRRDLEATLOHEATMAALRRKHADSAVLEGEIDNLOVYKCKLEK 1221
 QY 920 SKREEMVLDGLKMEATEVEKLFGRALPIRKMAVATHEFTSPCHDRIFESSDGIORLG 979
 Db 1222 SEKKELLD-----VTSNMQIITKANALEKM-----CRT--LEBOMNHRSS-----KAE 1264
 QY 980 ITRKRNGTFLKILPPIISADLRTGGRSSRPL--SAPRSGSKGFIIDVLPKSNEL 1036
 Db 1265 ETORSVNDL-----TRRAKLQTEGELSLODEKALLISOLTRC-----KILTYQOL 1312
 QY 1037 KETIAPLLIKLVGLACAFIIVPSADAVDAKTCACLLKGRIEFIACIANPACANVAC 1096
 Db 1313 EDLKRQLEEEVAKNALAHALQSAHDCDLLREQYEETETAKAELQRYLSK--ANSFVAQ 1370
 QY 1097 LOTCNRPDETCQJIKGDLFENSIVYDEFNECAVSRRKCVPRKSDLEFPAPDPBSVLVN 1156
 Db 1371 WET-----KIETDAIQRTEELEERKKKLAORIDABE----- 1403
 QY 1157 FNISDFNGKWYITSGINPTFADFQCLHEPHEEGDNKLVGNISWIRKITLDSGFTFRSAVQ 1216
 Db 1403 -----AVEAVNNAKCSLEKTRHRLONEIEDLMDV-----VERSNAA 1438
 QY 1217 KTVODPNOPGVLYHNHNEYLHYQDDWYILSKINKEPEDYIFVYRGRNDAMDGYGAVY 1276
 Db 1439 AALDKKO-----RNFDKILVEMKQYEESSQSELESQKE-----ARSL 1477
 QY 1277 YTRSSVLNLSIIPLE-----KAASIGRDFSTFITDNTGCGEPALVIRIKTYVEEG 1330
 Db 1478 STELFKLKNAYEBSLEHETFKRENKNQOEISOLTEOLSTGSIHELEKIKRQOLEAKR 1537
 QY 1331 RIIVKEVEIEEVEKEVEKGTENTLFO-----RLAEGNELKODEENVYRSL-- 1382
 Db 1538 LELQALBEAASLEHBEKGLTRAQOLEFNQKATIEKRLAKDEEMQAKRNHLRAVVDL 1597
 QY 1382 -----KEEME-FLDEIKWEASEVEKLFQKA 1405

Db 1598 QTSIDAETSRNEALRVKMKMGEDLNEMEIOLSHANMAABA 1639
 RESULT 12
 A:71928
 Cag island protein - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
 C:Accession: A71928
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
 ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
 A:Reference number: A71800; MUID:99120557
 A:Accession: A71928
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1819 <ARN>
 A:Cross-References: GB:AE001481; GB:AE001439; NID:g4155005; PID:g4155007
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: orf13/14 ;
 Query Match 1.9%; Score 142.5; DB 2; Length 1819;
 Best Local Similarity 19.0%; Pred. No. 1.1;
 Matches 267; Conservative 199; Mismatches 524; Indels 417; Gaps 68;
 QY 178 LFENSVQVDFNECAVSRRKCVPRKSDVGEPPVDRNAVONFMKDFSGKWTITSGINPT 237
 Db 244 LFSRSIFHYF-----VPLEDKSSRFS-KDRNLVND----- 274
 QY 238 FDFACOLHEFHENDKLVGNLTWRITKIDLGGEFTFRSAVOTFVQDDPLGALYHNHNEFL 297
 Db 274 -----EIQIROQ---EYNRILKERNKGMIDKNLE-----FNDNPN--RTLYN-- 312
 QY 298 HYQDDWYILSSQIENKPPDYIFVYRGRNDAMDGYGGSVLYTRSPITPESIIPILOKAK 357
 Db 312 -----YINIAEIDK--NPLRAFYECISN-----GSG--YECCKLIND--KKLODK 354
 QY 358 SVGRDFNPFITDNDSCGPEPVLREKTAEESEK--LIKAEVET-----EEEV 405
 Db 355 KTLNAVNDCI---KNATTEERIKCDLIDENILKKSILNQKVVALDCLNNAKTDEER 411
 QY 406 EKEVEKYVDTEM-TLFORLLEGFELQODEENFVRLSEER-ELINLOMATVEYKLF 463
 Db 412 KECCLKINDPEIRKFKFKELELOKELOEYKDCIKNAKTAENKECCLKGSKEA--IERLK 469
 QY 464 GRALPIRKLRLMALPHSNFIANHEITIKYVGSKLPGHKKRFSWGMEDYFSGIYVAKICSR 523
 Db 470 QQALDCLKNNAKTDEERCKLN--IPDOLQKRLADM-----SVAYADCVSR 515
 QY 524 RIRYIFK-----SPKICGLDSRGJOLFSGHKNLSPAHSINONVVK-----GNSG 570
 Db 516 ARNEKEKQCEKILLTPPAKKLLENOALDCLNNAKTDEERCKL-KNLPLODLSILAKES 574
 QY 571 CKFPKDYA-----LAWKMGQCFAKIAYAFILTSASADAVDAKTCCTC 616
 Db 575 LKAYKDCASQAKTDAEKKCEKILLTP-----AKKILEEAKESVAYLDCVSOAKT-EA 628
 QY 617 LKKEC-----RLEIAKISNPACANVACLOTQNNRPDETCO-----IKGDLFE 662
 Db 629 EKKECEKILTPPAKKLLEAK-----KSVRAYIDCVSKANANEKRECKLILTPPAKKLLE 684
 QY 663 NSVVDENECVARRKCVPRKSDVGDPPVDPBSVLVQKEDMKDFSGKWFITRGINPTDA 722
 Db 685 NOALD-----CLKNAKTDEERCKLNDLP-----KDLQKYIAKESVAVYIDC 727
 QY 723 FDCOLHEFHTEENKLVGNLSWIRTPPDGFTFRSAVQKRVQDDPKPIGLY----- 773
 Db 728 VSKAKNAEAKKEC-----KILTIPE-----ARKLLEAKKSVRAYKDCVLRARNE 772

surface-located membrane protein Imp3 precursor - Mycoplasma hominis (SGC3)
 C:Species: Mycoplasma hominis
 C>Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 05-Dec-1998
 C:Accession: J06009
 R:Adapted, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
 J. Bacteriol. 178, 2775-2784, 1996
 A>Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis Imp gene system
 A:Reference number: J06009; MUID:96212016
 A:Accession: J06009
 A:Molecule type: DNA
 A:Residues: 1-1302 <LAD>
 A:Cross-References: EMBL:X95601; NID:g1197335; PID:e222344; PID:g1197336
 C:Genetics:
 A:Gene: Imp3
 A:Genetic code: SGC3
 C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homology
 C:Keywords: duplication; membrane protein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>
 F:957-1026/Domain: tetratricopeptide repeat homology <TR1>
 F:993-1026/Domain: tetratricopeptide repeat homology <TR2>
 F:1089-1120/Domain: tetratricopeptide repeat homology <TR3>
 F:1154-1190/Domain: tetratricopeptide repeat homology <TR4>

Query Match 1.9% Score 140.5; DB 2; Length 1302;
 Best Local Similarity 18.7%; Pred. No. 0.92;
 Matches 173; Conservative 123; Mismatches 312; Indels 315; Gaps 40;

```

Oy 145 KCIAPRS-----CAANVACLOT-----CNNRPETCCQIK--CGDLEF--NSVDQONE 189
Db 445 KSVINSSKSKSDIIANENLIDALADANKAKDQVDEAKSKISKEQJLMLIKANTLLPOLND 504
Oy 190 CAVSRRKCVPRKSDYGEFPPVDPNAVQNFNMKDFSGKWTITGLVPTDAPDCOLHEFH 249
Db 505 ---NDEIVKAKESLNEITNANKAVNONDNASQSK---SLDDKTKTKIONQLTEFN 557
Oy 250 MENKIKVNLTKRIKTLDDGFFTRSAVQTVQDPDLPGLYHNDNEFLHYODDWYLLSSQ 309
Db 558 KDKNA-----KRELEQ---TRKIDNFLIDP-----YKNPN-----YATLVAD 594
Oy 310 IENRPDIYFYVYRGRDAMDGYGGSVYTRSPTEPESILPN-----LQ 353
Db 595 LTNKDKKSY-----TKSSNKSEIITANDELKQALDKKAVAKDQD 636
Oy 354 KAASVGRDNRNFTTIDNSCGPEPLVERLEKTAEGEKLKKEAVEIEEVEKEVEKVR 413
Db 637 EANKSIEKQLSDSITNANO-----LNLKL---VDSKDIOKAKTELSQELQSASQELN 686
Oy 414 -----DTEMTLFORLLEGFKELQODEENEFRELSEKEKELINLEOMEATEV 459
Db 687 LNNPTSMOSAKESLDKAVTEITKLEF---NKDKVKEFELEKTRKD----- 733
Oy 460 EKLFGALPIRLMLAPHNSFLANHEITKYVVGSKLPGHKRFSGWGEDYFGSIVAKI 519
Db 733 -----DEFINTNKNPN-----STLISEL 752
Oy 520 CSSRRIPRYRKRPRICCGDLSRGDLFSGKHNLSPAHS-INQVPEKNSGCKFPKDYA 578
Db 753 TSKR-----DSKNSVTNSNKSQDIETANTTELKQALAKANTDKAQNLA 796
Oy 579 LMMWKGCFAKTAIFAIFILSVASAKADAVDLTKTCTCLKKCRLELAKKCSNPACAA-N 637
Db 797 KSTKEQLNNSISSA--NTLLAKTLKDNITIOAKT-----ELEVEVK--ANAQVAVSN 846
Oy 638 VALQICNNRPDETCQI-KCGDLENSVDEFECAVSRKKCVPRKSDVDF-----PV 691
Db 847 TASMOAKSSLDKAVTEITKLEFKNKDKVFKLEQGRK-----DIDEIFINTKTN 899
Oy 692 PDPSVAVQFDMKDFSGKWFITRGLNPT-----PDADC----- 726
Db 900 PDSTLISELTSKRS--KNSITNSNKSQDIETANTTELKQALAKANTDKAQNLAARSTKE 958

```

```

Oy 726 QLHEFTPEENKLVNLSMRIRTPDGGFFTRSAVQKVVQ-----D 764
Db 959 QUNSSISANTLLAKLKDNDKNTIOQ---AKTELEKEQYANOAVASNNTASQSKSSLD 1015
Oy 765 PKYPCI-----LYHND-----EYLLYODDWYLLSKVE-----NSPEYIFV 802
Db 1016 AKVTEITKLETFENKDKDQVDFRELEQGRKIDDEFINTNKNPNYSTLISELSKRD----- 1072
Oy 803 YKGRNAMDYGGSVLYTRSAPVPSII-----PELQTAQKGRDPNFTIKYDNTCG 856
Db 1072 ---SKNSTTNSNKSQDIETANTTELKQALAKANTDKAQNLAARSTKEQLNKSISANT-- 1127
Oy 857 PEPLVERLEKKEVEGERITIKVEEIEEVEKVR-----DKETVLF 898
Db 1127 -----LAKTLKDNITIOQAKTELEKEQYANOAVASNNTASQSKSSLDIKVTEI 1178
Oy 899 SKLEFGKELQRODEENLRSLK 921
Db 1179 TKKLEF---NKDKAKFENELK 1198

```

RESULT 15

F64508

hypothetical protein MJ1672 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997
 C:Accession: F64508

R:Bult, C.J.; Overbeek, R.; Kikunas, E.F.; Weinstock, K.G.; Merrick, J.M.; Glöck, J.; Reich, C.J.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
 A>Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-800 <BUT>
 A:Cross-References: GB:U67607; GB:L77117; NID:g1592244; PID:g1500575; TIGR:MJ1672
 C:Genetics:
 A:Map position: REV1656207-1653805

Query Match 1.9% Score 140; DB 2; Length 800;
 Best Local Similarity 20.0%; Pred. No. 0.49;
 Matches 180; Conservative 96; Mismatches 228; Indels 394; Gaps 45;

```

Oy 650 ETEQIKCGDLFENSVDENECVSRKKCVPRKSDVGFPPDPV--LYQKF--DMKD 705
Db 119 ETVC---IGELTENLYKNGFK---YSKLPLVSDAIFTRKPRYENKELFSFEDEIKD 172
Oy 706 FSG-----KWF-----TGLNPTDAED-----COLHEFH 731
Db 173 FKGDNVFEELVQLMQKWTWCIPSVTMMKAGSLKGGPVDLSFHSKTCATCAICLYQWY 232
Oy 732 TEENKLVNLSMRIRTPDGGFFTRSAVQKFEVDKPYGILYHNDNE-----LLV----- 782
Db 233 VKENK-----KKNKTAKEYIDKYTELEKLPNDNGNKKIFSLINGDLS 275
Oy 782 ---QDDWYLLSKVENSPEDYIFYVYKGRNDAMD-----GYGGS 817
Db 276 GIQDFVFTTK-----YATKSLKGRSFYLDLTFEFAKYICKELNLPITNLRYGGG 328
Oy 818 VLYTRSAVLPSIIPETLOTAQKGRDENE-----ITDWTGCEPPEPLVERLEKKEVEE 873
Db 329 HNYLLSYVDENLNLKLEKLENEY--LENNFRIKYITTAIEVAVTNDPKKSEDK--ESKE 385
Oy 874 RT--IIEVEEIEEVEKVRDKETLFSKLFEG--FKELQDEEN----- 915
Db 386 KIWGFAKKKKEVS--EKVTEKKLRREFYKLEGLFEPYNNGSENRVCYICNEDKNEKGY 442
Oy 915 FLRE--LSKE-----MDVLDGLMEATEVEKELFGALPIRLKLMATAVTHCTSP 961

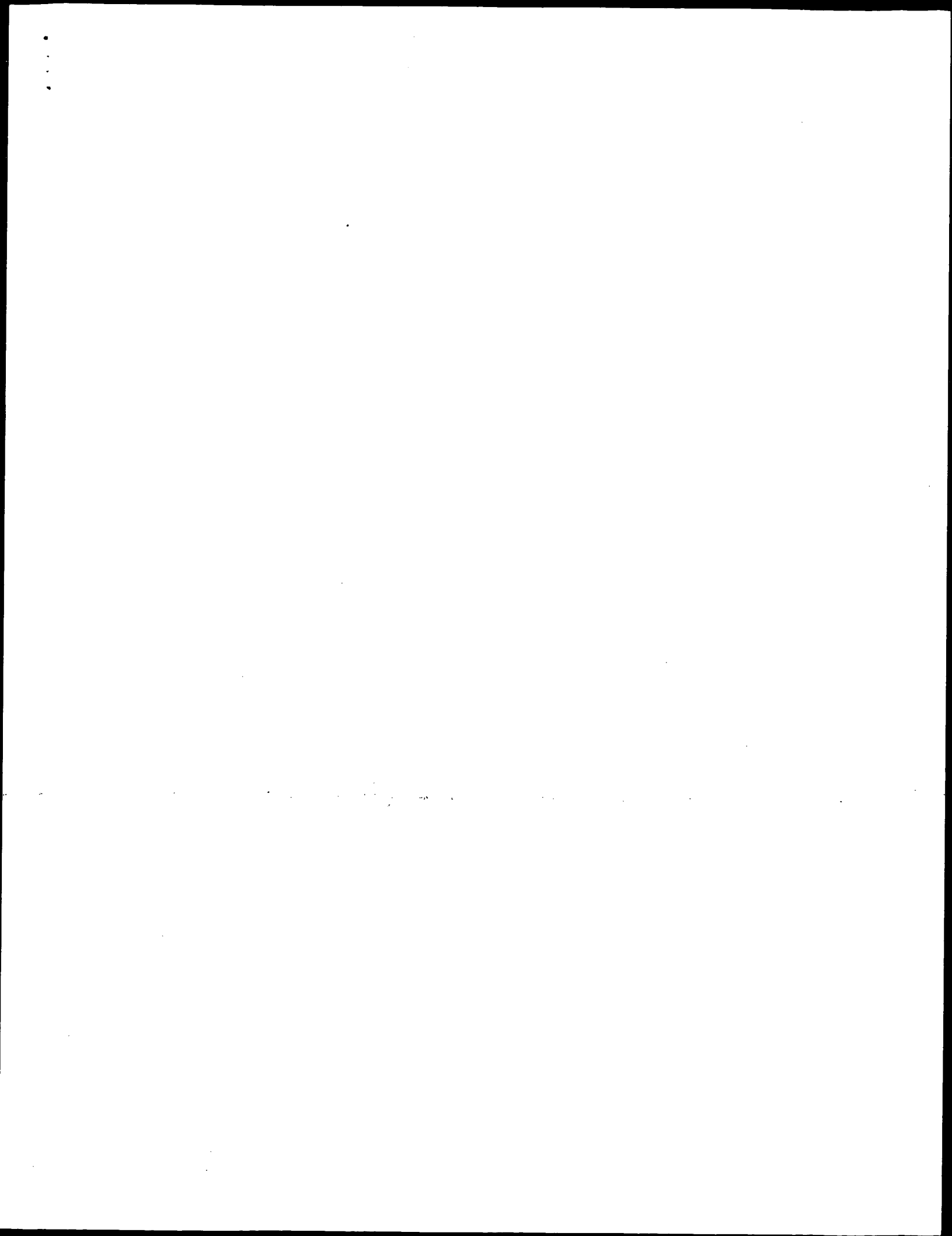
```

```

Db 443 AIRENESKERIDYCASFAVLTDLKKNFOMEKT---IKFNKAVPI----- 486
QY 962 CHDKIRFFSSDDGIRGIRKIRKIRNGTFLKILPIIOSADLRTGGRSSRPLAFSGFS 1021
Db 486 -----: : : : :-----LLOREES 503
QY 1022 KGIDIVLPBKNELKELIAPLLKLVGLACAPLIYPSADAVDALKTACALLKGCRIEL 1081
Db 504 -----: : : : :-----PLTV---KALEKLESKFRVLSD----- 523
QY 1082 AKCIANPACANVACLOTNNRPDETECQIKGDLFENSVADEFNECAVSRKCVPRKSD 1141
Db 523 -----: : : : :-----ENYFLKEKILPHDSGELIIPYKIW 546
QY 1142 LGEPAPDPVLYON---FNISDFNG---KWIYTSG-----LNPTDADFQQLHEFHTE 1189
Db 547 AIAFP-----: : : : :-----IENETEKRIIDFDGLAEKAFERTGTIRKIGILKMDVD---NLGEIETT 596
QY 1190 GDNKLVGN---ISWRIKTLDSG---FTRSAVQKFVODPNQGVLYNHNDEYLYHQQDWY 1243
Db 597 G-----LGNDATIS--RMSLSSMLTLEFTGYI-----: : : : :-----PHLIKNEEFE----- 633
QY 1244 ILSKTIENKPEDYLFVYIRGRND-----AMDYGGAVVYTRSSVLPNSIIPLELKAAXSI 1298
Db 633 -VNGK-KYKFKDNIIYLYAGGDDTLIVGAMDA-----: : : : :-----VWELAKRI 670
QY 1299 GRDSTIFRTDNTC-GPEPAL-----: : : : :-----VERIEKTVEGERTIVKEVEEIE 1341
Db 671 RGDEFKFFV-----CYNPYITLSAGIVFVNPKEFEKKAANMAEELLENGKNYIITYEDEETE 725
QY 1342 EEVEREVEKVGRTFM-----: : : : :-----TLFORLAEGFNEKODENENFRELSKEME 1386
Db 726 KKVDPKNALIVNCPNMDLEVYENYCWTKRLKSYLEGINKEMVELLESIVKKKNEDDLE 783

```

Search completed: October 14, 1999, 03:19:51
 Job time: 2237 sec



GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 1999, 03:57:50 ; Search time 26.17 Seconds
(without alignments)
1525.213 Million cell updates/sec

Title: US-09-075-375-6
Perfect score: 7495
Sequence: 1 MALSLHTVLCKEELNLYA.....MEASEVEKLEFGKALPIRKVR 1412

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	158.5	2.1	1957	1	YD86_SCHPO
2	155.5	2.1	2869	1	RBPL_PIAYB
3	155	2.1	2469	1	TEG0_HSYA
4	148.5	2.0	1935	1	MYSB_PIG
5	146	1.9	1935	1	MYSB_HUMAN
6	143	1.9	2748	1	NDM1_YEAST
7	142.5	1.9	1935	1	MYSB_RAT
8	137.5	1.8	1875	1	MLP1_YEAST
9	135	1.8	3660	1	DMD_CHICK
10	134.5	1.8	1937	1	MYSP_HUMAN
11	134	1.8	1630	1	MSP1_PLAFK
12	134	1.8	1639	1	MSP1_PLAFK
13	133.5	1.8	1934	1	MYSB_MESAV
14	132	1.8	809	1	LEF_BACAN
15	130	1.7	1679	1	XIO9_YEAST
16	129.5	1.7	1805	1	HMM2_MYCGE
17	129	1.7	1822	1	YMO8_CAEEL
18	127.5	1.7	1790	1	USO1_YEAST
19	126.5	1.7	4568	1	DYHB_CHLRE
20	126	1.7	993	1	NI8B_LACLA
21	126	1.7	1744	1	TANA_XENLA
22	124	1.7	2284	1	POL1_SPLV
23	123	1.6	3678	1	DMD_MOUSE
24	122.5	1.6	1341	1	RPAL_METJA
25	121.5	1.6	1039	1	SYI_METJA
26	121.5	1.6	1545	1	YH2_YEAST
27	121	1.6	2025	1	TTG3_HUMAN
28	120.5	1.6	2418	1	SPCA_HUMAN
29	120	1.6	610	1	CALG_HUMAN
30	120	1.6	1324	1	CUT3_SCHPO
31	120	1.6	4725	1	DYHC_DICDI
32	119	1.6	881	1	ARPB_YEAST
33	119	1.6	2411	1	MYSA_DROME
34	119	1.6	709	1	TOP1_MYCGE
35	119	1.6	1102	1	YME6_YEAST
36	118	1.6	2663	1	CENE_HUMAN
37	117.5	1.6	857	1	PIFI_YEAST
38	117.5	1.6	1230	1	UGS4_SOLTU
39	117	1.6	2093	1	RRPL_TOSV
40	117	1.6	1940	1	SPCO_MOUSE
41	116.5	1.6	1208	1	MYSE_RAT
42	116.5	1.6	1208	1	YD86_SCHPO
43	116.5	1.6	823	1	YHX2_YEAST

ALIGNMENTS

RESULT	ID	YD86_SCHPO	STANDARD	PRT: 1957 AA.
44	116	1.5	645	1
45	116	1.5	1302	1
				RRPL_BTV10
				P26413 glycine max
				P13840 blue tongue

Query Match 2.1%; Score 158.5; DB 1; Length 1957;
Best Local Similarity 17.6%; Pred. No. 0.24;
Matches 265; Conservative 227; Mismatches 503; Indels 511; Gaps 68;

QY	59	YKTSFSDSSHCKDKSQICISIDTSFEIQRFDKRGKTLLEKQMFQIOLAVLVCTEV	118
DB	305	HNVANYSDAIYHKDK-LIEDISTRISFDMKSEBDTLSTKNEKTLKLR-----	354
QY	119	IVPRYDAVDALF--TCACLLKEGRITLACIANPSCAAVACLOTGNNRPDECOIKC	175
DB	354	-----NTIGSLKDSRTSNSOLEEVEVLEK-----SN-----RTIHSQTLTAESKLS	396
QY	176	GDLFENSV--VDFNECAVSRKCKVPRKSD-----VGEFPVDRNAVYQNFN	220
DB	397	FEOENKSLKSGIDEXONKLSKDKMVKQVSSOLEEANSLSAHATGKLAIEINSEBDQNK	456
QY	221	MKDFSGKWTYSGLNPTFDACOLHEFH--ENDKLVGNLTKRITLDG-GEFTRSAV	276
DB	457	IKDFE---KIFCDLRACILNSSNELKERSALIDKDELNLNRLQIKQKVSSESTSSL	513
QY	277	QTFVDDPLPLGLVNHDEFLHYDDWVTLSSQIENKPPDIIFYRYGRDAMDGYGSV	336
DB	514	QSLQD-----ILNEKKKHEVYSQLELKEGLQTE-----	545
QY	337	IYTRSPILPES;TNLOKRAKSVGRDNFNFTTDSNGCPEPPLVERLEKTAEGEKLILK	396
DB	545	-----INSEHLSQSLSTAELAEKAAVATNNE-----LSEKNSLQOT	581
QY	397	EAVLEIEVEKVEKVRTEM-----TLPRLLGKRELOODEENVRLESEKELILN	450
DB	582	LCNAQOEKRLAKSVMLKENEQNFSLDTSFKRLNESQOELENNHQTITKOL-KDTSSKLO	640
QY	451	ELQMEATEVEK-----LFGRALPIRKLMALAPH-----SNFLANHEIKIYVGS	495

```

Db 641 QLOLERANFEOKESTLSDENNDRITLKLKLEESNKSLLIKKQEDVDLEKNTQITLK----- 696
OY 496 KLGHKRFSGWGEDYFGSIYVAKICSSRIPIYFRKSPICCGDLSRGLQLEF---SHGHK 552
Db 696 -----ED-----LRKSEDLRFKSK-----LEAKKLREVIDNLKSKH 726
OY 553 NLSPA-----HSI-----NONPKGNSGCKFPKDVALLWMEKMGQFATAIYAFIITISYA 602
Db 727 ETLERANDHSLSDAKNTNAILSELKSSSEV-----KRLANVEITLQD 774
OY 603 SKA-----DAVDALKTCTCLLKECHLEIACISNPACANVACLQTCNNRDEFEQIK 656
Db 775 SKAMKOSFTSLVNSYOSISNLYHELRDHDVN-----MOSQNTLLESESKLK 821
OY 657 ---CGDLFEN---SVDEF-----NECAVRKKCVPRK-----SVGSGFPVPDP 694
Db 822 TDCEENTLQONMTLIDVQKLMKHNQESKYSSELKEVNGKLSLDKRLSSLVNAISDND 881
OY 695 SVLVQKFDW-KDPSGKWFITRGILNPTDPAFCOLHEPTEE-----NKLVGNLVWRIR 746
Db 882 QILTQLAELSKNYSDLEQESAOINSGLSLEAKQQLHTEMBELHRLDKITGKTK----- 938
OY 747 TPDGGFTSAVQKFOVPKPYGIIYNNDNEYLYQDDWYLLSSKVENSPEDYIFYVYK 806
Db 938 -----IEBSKSSDL-----GKKILARQEE---ISNKEENMSOSQALITSVKS 976
OY 807 RNDANDYGGSVLYTRSAVLPESILPELOTAACKVGRDFNTFIKTNDGCEPREPLVERLE 866
Db 977 KLDE-----TLSSKSKL-EADIEHLKKNVSEVEVERNMLLASN-----ERLM 1017
OY 867 KAVE-EGERTTIKEVELEBEVEKYRQKEVYLFESKL-----FEGF-----KELORD 911
Db 1018 DDLKNGE-----NLSAQTEIEKKRAENDLOSKLSVSSSEYENLLISSQTNKSH-E 1071
OY 912 EBNFLRELSEKEMVDLDGKMEATEVEKL---FCR-----ALPIRLKMAVAATHC 957
Db 1072 KLNQKYLEKKNVQKLLDKDKNVLEBILTSKYKLEENNAQIKDELALAKRK----- 1125
OY 958 FTSPCHDRIREFSSD-----DGIGRL-----GITRRKINGTFLKILPIQOS--AD 1001
Db 1125 -SKRQHDICANFVDLKEKSDALEQLTNEKELLIVSLQESNSNMEALVERPSLANRLSD 1183
OY 1002 LRTTGRRSRPLSAFRSGFSKIGDIYPLPSKNELKLTAPLLKLVGLACAFLLVPSA 1061
Db 1184 MKKSLSDSDNVIVIRSD-----LVRYN 1206
OY 1062 DAVALDKTCACLLKGCRIELAKCIANPACANVACLQTCNNRDEFEQIKCGDLFENSV 1121
Db 1207 DELDITLK-----KDKDLSLTQYSEVQCDRDDLLDSLKGC----- 1241
OY 1122 VDENECAVS-RKKCVPRKSDIGEFPAIDPSVLYQN--FNISDFNGKRWYT--SGLNPTFD 1177
Db 1241 EESNKTAIVSLRELCTSEIDV-----PVSEIIDDNEVFAGNFSLSRLTYLSLENYLD 1295
OY 1178 AFDCLAEFHEGDKLVGNISWRIKTLDSGFFTRSAVOFQVDDPQVGLVHNHNEYH 1237
Db 1296 AFN-QVNFKKMEIDN-----RLTTTDAEF-----TKVAVDL-----EKLOH 1330
OY 1238 YODWYLLSKIKENKPPEDYIFYVYRGNDAMDYGGAVVYTRSSVLPNSIILELEKAAS 1297
Db 1331 EHDWML-----QRG-----DLKALKD 1348
OY 1298 IGRDFSTFIRTDNCGEPALVERIKTYEGERIIVKEVELEBEVE----- 1346
Db 1349 SEKNF-----LKRKAEAMTNIH-SLEEGKEETFKKIELAESRLEBNOATNKLKNO 1398
OY 1346 -----KEV-----EYVGRTEMLTFORLAEGFENELKODENEFVRELSEKEMFEL--DEIKKEA 1395
Db 1399 IDHNLNOEIRLKEVDLKEKESLISLSESLSNQOKOKSSSLD--ANNELEHMLDITSRKNS 1456
OY 1396 SEVEKL 1401
Db 1457 SLMEKI 1462

```

```

RESULT 2
RBP1_PLAVB
ID RBP1_PLAVB STANDARD: PRT; 2869 AA.
AC 000798;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1.
OS PLASMODIUM VIVAX (STRAIN BELEM).
OC EURAROTIA, ALVEOLATA, APICOMPLEXA, HAEMOSPORIDA, PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92315338.
RA GALINSKI M.R., MEDINA C.C., INGRAVALLO P., BARNWELL J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL CELL 69:1213-1226(1992).
CC -I- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -I- SUBUNIT: HOMODIMER (POTENTIAL).
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M88097; G160626; -.
DR HSSP; P36956; IAM9.
KW MALARIA; RECEPTOR; SIGNAL; TRANSMEMBRANE.
FT SIGNAL 1 17
FT CHAIN 18 2869
FT DOMAIN 18 2807
FT TRANSMEM 2808 2826
FT DOMAIN 2827 2869
FT SITE 1030 1032
FT SITE 2599 2601
SQ SEQUENCE 2869 AA; 330213 MW; 2BD43E2C CRC32;

Query Match 2.1%; Score 155.5; DB 1; Length 2869;
Best Local Similarity 17.4%; Pred. No. 0.59;
Matches 283; Conservative 242; Mismatches 573; Indels 529; Gaps 76;

OY 57 TSKRTSFSOSSHCKDQSGICSDTSPFEI-----QRFDLKRGMTLLLEQWQFIOLA 110
Db 1150 TSNKNELLKSVYEEVEKLVVQONEDYKKNVKNENKQLEAIRSKSLKE----- 1201
OY 111 IVLVCFVIVPRYDAVDALKTACLLKRCRIELAKCIANPACANVACLQTCNNRDETE 1170
Db 1201 -----VINKHSEWTOLESTANTLK-----SNAGKRNEDLEELN--KTK 1239
OY 171 COIKGDLFE--NSVVDQENECVSRKKCVPRKSVGGEFPVPDPA--VVQNFNM-KDF 224
Db 1240 GQMR--DIYEKLKRIAEELKEGTVELNDANEKANKE-PEPERNIIGHYLERITVEKDK 1296
OY 225 SGKWTYSGLNTPDPAFCOLHEF--HMEKDVLGNLTWRKTLIDGGEFFTRSAVQTFVQD 282
Db 1297 AGR--VVEEMNSLKRIKILQLEISDSDQNLVTTSTIKHLNAGK----- 1341
OY 283 PDLPGALVNHNEFLHYDDWYLLSSQIENKPPDYIFYVYRGNRN-DAWDYGGSVLYTR 341
Db 1341 -----YED-----VIKRNEDSIQLEKAKSLFELD--EM 1368
OY 342 PTLPESTIIPNLOKAK--SVGRDFNN-----FITDNS----- 373
Db 1369 KTLVQOVNMNLOSALQGNAGISKELNELKGYITELLISTNYSILEYVKKNSSESVRSQSL 1428

```

```

QY 373 CGPEPLVERLEKTA-----EGEKILLIKEY-----EIEEVEK-----EVEKVRDT 415
Db 1429 ANGFTKAEGEKNASARLAERKL--KEQIVKDLVDYSDIDDKVKKLEGIRKREILKMKES 1486
QY 416 EMTLFORLEGEKQOQDEBNFVRLSKKEEILN-----ELOME----- 456
Db 1487 ALTFEEF--SEKFKQMCSSHMEKAKKIEYLKNNNGGKANITDSDMEVGVYVSKAE 1545
QY 456 -----ATEVEKLFGRALPIRKLMALAPHSNLANHETIKY 492
Db 1546 HAFHTEVAQVKTAFCEISIVAYVTKMDNLEWESL--MKEVKCKEK-----KNEAEKRY- 1599
QY 493 VGSKLPGHKRFSWGWEDYFGSIVAKICSSRIPIRYFKRSRICOGLDROLOJFSGKH 552
Db 1599 -SAKL-----KPYDGR1-KARVSENERKISELEKAKV--EKKESSOLDVSTK 1643
QY 553 NUSPAHSINQVPKGNSGCKFPKDVALLWMEKWOFAKTAIAPILSVASKADAVDAK 612
Db 1644 SLLOJDNCRQOLDVSLNIGRYKQNALOYFDS---ADKSKMSVLPISIELGAESLDAKVK 1699
QY 613 TCTCLLEKCERLACISNPACANV--ACLOTCNNRDETECO--IKGGLFENSVDDEF 669
Db 1700 AA-----KESYEKNELETVQEMSRINVEEGSLTDIDKIIDENDLKKKKOYEGLDOKI 1755
QY 670 NECAVSRRKCVPR--KSDVGDFFVPPPSVLVQKFDKDKSGKWFITRGILNPTFADACOLH 728
Db 1756 KENADKRKSNFELVGESEINALLDPSTISFI-KLKLKEYD-----MGDLKKNYGVKKN 1806
QY 729 EPHTEENKLVGNLSMR1-----RTPDGGEFFTRSAVQK 760
Db 1807 EIHGEFTKSYNLIEHTLSNATDYVTFEKAQSLRELAKEEHLRREBEALFLLNDIKK 1866
QY 761 FVO-----DPKYGILYHNDNEYLLYQDDWYIIS----- 790
Db 1867 VESLKLKEMMKVSAEYEGMKRDHTSVQOLVQDMKTIYDELKTLNDISECCSVINNVYS 1926
QY 790 --SKVENSPE-DYIFVYRGNDAMDYGSGSVLYTRSAVLPESIIPELOTAOKVGRDEN 846
Db 1927 IYKVKESKHADY-----RSDANSMTYSMTLANFISDEKAKISS-----GMEFN 1971
QY 847 TFIKTD-----NTGPEPLVERLE-----KVEEGER-----TIIKEV 880
Db 1972 AEKMSKFTDLEIFISVSNSENLKITEQDSDVYQKRESEQJAKDATDIYVNIKLK 2031
QY 881 EETEEVEKVRDEKVTLLFSKLFEGFEKLODE-----ENFLRELSEKEMDVLGKAKMA 934
Db 2032 NEFNEKLEEKKEEYVSEKVRREALKRLSQVEGIRCHFENFRLDNT----- 2081
QY 935 TEVEKLFGRALPIRKLMVAVTHCFTSPCHDRIRFFSSDDIGRLITRRKINGTFLKIL 994
Db 2081 -EIEENL-----KKWYTI-----RDKKSEFESLOEM--ENEMAN-TYNSNT- 2118
QY 995 PPIQSDLRITGGRSSRNPISAFRSFGSKGIFDIYVLPKSKNE--IKELTAPILLKLVGV 1050
Db 2118 --TQLESTIVASGSK-----DIEKLEERSNEEMRINISEKISTIDSK----- 2158
QY 1051 LACAFILVPSADAVDAKTCACLLKGRIFLACIANPACAAVAVACLOTCNNRDETECO 1110
Db 2158 -----VIEKNSITIDELYK--LQKNQAHWISLISITANMKTSKILMINKEKENTE- 2207
QY 1111 IKGGLDE--NSVVDENECVAVSRKRCVPRKSDIG--EPAPDPSPVLVONENISDFNGKWV 1167
Db 2207 -KCYDYIKDNSSSTDGYVELT-----KGFYGSKLITSSASEIYQANDDYVSNFAKH 2256
QY 1168 ITSGILNPTFAFDCQLHEFTEGDNKLVGNISWRIKTLDGFTTRSAVQKVFVODPNQGV 1227
Db 2257 EKESILNLRD--IKKELYLFHONSISISIV-----EGGVQNMU-----A 2292
QY 1228 LYNNDNELIYQDDWYIISLSEKPEPDY-----IFVYRGNDN----- 1268
Db 2293 LYDKLNEKREMDIELYRINISETKLQMEHSTIDVCKPMIELHKGNNETNNKSLEKEKKL 2352

```

```

QY 1268 -----WDYGGAVYTRRSV--LPN--SIPELEKAKSIGRD----- 1302
Db 2353 SVNDHMSAEKIRKGLTYTPESVQNNINIVSIEAEKTLDEIDRPGDNYQIVEBK 2412
QY 1302 --FSTFRDNTNCGPEPALVERIEKTEGE--RIIYKEVEEIEEVEKVEV--GR 1353
Db 2413 KQPSILIDRTN-----ALMDIEIFKKNYNLMENVNTEIHRVNDYETIKTNLVQAK 2466
QY 1354 TE-----MTLFOR--LAEGFNEKODEBNFVRLSKKEE-----EPDLE 1390
Db 2467 TEYEQLENTKQNDMDLQNLFLKYSIIYFENVKKKESILNDLYEQDRLLKIEHDE 2526
QY 1391 IKMEASE 1397
Db 2527 IKRNATE 2533

RESULT 3
TEGU_HSVSA STANDARD; PRT: 2469 AA.
ID TEGU_HSVSA 001056;
AC 001056;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE PROBABLE LARGE TEGUMENT PROTEIN.
GN 64 OR EEF2.
OS HERPESVIRUS SAIMIRI (STRAIN 11).
OC VIRUSES: DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92333688.
RA ALBRECHT J.-C., NICHOLAS J., BILLER D., CAMERON K.R., BIESINGER B.,
RA NEWMAN C., WITTMANN S., CRAXTON M.A., COLEMAN H., FLECKENSTEIN B.,
RA HONESS R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. VIROL. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92230228.
RA NICHOLAS J., CAMERON K.R., COLEMAN H., NEWMAN C., HONESS R.W.;
RT "Analysis of nucleotide sequence of the rightmost 4.3 kbp of
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT organization between HVS and Epstein-Barr virus.";
RL VIROLOGY 188:296-310(1992).
CC CC -1- FUNCTION: TEGUMENT PROTEIN.
CC CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC CC EHV-1 24, EBV BPFL1, HVS-1 64, VZV 22, AND HCMV UL48.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: X64346; G60385;
DR EMBL: M86409; G330993;
DR PIR: H36812; H36812.
SQ SEQUENCE 2469 AA; 280165 MW; 1B72F5F9 CRC32;

Query Match 2.1%; Score 155; DB 1; Length 2469;
Best Local Similarity 17.3%; Pred. No. 0.51;
Matches 230; Conservative 208; Mismatches 497; Indels 396; Gaps 58;

QY 271 FTRSAVQTFVQDPLPGALYNHNEFLHYQDDWYIISQINRKDDYFYVYR----- 324
Db 178 FSTSVNDAIEYLSPPNVOYT--GSFLYFVPEYIGHSH-----YIMNHYRVINTEKL 228
QY 324 -GRNAMDYGGGSVY-----TRSPLEPESIIPNLOKAKSVGRDPNNFI 367

```


RA MEDLINE; 86176778.
 RA SAEZ L., LEINWAND L.A.;
 RT "Characterization of diverse forms of myosin heavy chain expressed in
 RT adult human skeletal muscle.";
 RL NUCLEIC ACIDS RES. 14:2951-2969(1986).
 RP REVISIONS.
 RP LEINWAND L.A.;
 RL SUBMITTED (MAR-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [17]
 RP SEQUENCE OF 1410-1935 FROM N.A.
 RX MEDLINE; 88299163.
 RA KURABAASHI M., TSUCHIMOTOCHI H., KOMURO I., TAKAKU F., YAZAKI Y.;
 RT "Molecular cloning and characterization of human cardiac alpha-
 RT beta-form myosin heavy chain complementary DNA clones. Regulation of
 RT expression during development and pressure overload in human
 RT atrium.";
 RL J. CLIN. INVEST. 82:524-531(1988).
 RN [18]
 RP SEQUENCE OF 785-1935 FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RX MEDLINE; 90235862.
 RA BOBER E., BUCHENBERGER-SEIDL A., BRAUN T., SINGH S., GOEDDE H.W.,
 RA ARNOLD H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL EUR. J. BIOCHEM. 189:55-65(1990).
 RN [19]
 RP SEQUENCE OF 1393-1935 FROM N.A.
 RX MEDLINE; 87192738.
 RA JANDRESKI M.A., LIEW C.-C.;
 RT "Construction of a human ventricular cDNA library and
 RT characterization of a beta myosin heavy chain cDNA clone.";
 RL HUM. GENET. 76:47-53(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE; 96039076.
 RA ARAI S., MATSUOKA R., HIRAYAMA K., SUKURAI H., TAMURA M., OZAWA T.,
 RA KIMURA M., IMAMURA S.-I., FURUTANI Y., JOH-O K., KAWANA M., TAKAO A.,
 RA HOSODA S., MOMMA K.;
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
 RT hypertrophic cardiomyopathy.";
 RL AM. J. MED. GENET. 58:267-276(1995).
 RN [11]
 RP VARIANTS HCM GLU-256 AND ARG-741.
 RX MEDLINE; 93248216.
 RA FANANAPAZIR L., DALAKAS M.C., CYRAN F., COHN G., EPSTEIN N.D.;
 RT "Missense mutations in the beta-myosin heavy-chain gene cause central
 RT core disease in hypertrophic cardiomyopathy.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 90:3993-3997(1993).
 RN [12]
 RP VARIANT HCM GLN-403.
 RX MEDLINE; 90367131.
 RA GEISTERFER-LORANCE A.A.T., KASS S., TANIGAWA G., VOSBERG H.-P.,
 RA MCKENNA W., SEIDMAN C.E., SEIDMAN J.G.;
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta
 RT cardiac myosin heavy chain gene missense mutation.";
 RL CELL 62:999-1006(1990).
 RN [13]
 RP VARIANTS HCM GLN-249; GLN-403; CYS-453 AND MET-606.
 RX MEDLINE; 92204193.
 RA WATKINS H., ROSENZWEIG A., HANG D.S., LEVI T., MCKENNA W.,
 RA SEIDMAN C.E., SEIDMAN J.G.;
 RT "Characteristics and prognostic implications of myosin missense
 RT mutations in familial hypertrophic cardiomyopathy.";
 RL NEM ENGL. J. MED. 326:1108-1114(1992).
 RN [14]
 RP VARIANTS HCM GLN-403; CYS-453; ARG-584 AND MET-606.
 RX MEDLINE; 94070863.
 RA WATKINS H., THIERFELDER L., ANAN R., JARCHO J., MATSUMORI A.,
 RA MCKENNA W., SEIDMAN J.G., SEIDMAN C.E.;
 RT "Independent origin of identical beta cardiac myosin heavy-chain
 RT mutations in hypertrophic cardiomyopathy.";
 RL AM. J. HUM. GENET. 53:1180-1185(1993).
 RN [15]
 RP VARIANTS HCM GLN-403 AND VAL-908.
 RX MEDLINE; 92346810.
 RA EPSTEIN N.D., COHN G.M., CYRAN F., FANANAPAZIR L.;
 RT "Differences in clinical expression of hypertrophic cardiomyopathy
 RT associated with two distinct mutations in the beta-myosin heavy chain
 RT gene. A 908Leu-->Val mutation and a 403Arg-->Gln mutation.";
 RL CIRCULATION 86:345-352(1992).
 RN [16]
 RP VARIANTS HCM LEU-403 AND TRP-403.
 RX MEDLINE; 94075629.
 RA DAUSSE E., KOMAJDA M., FETTER L., DUBOURG O., DUFOUR C., CARRIER L.,
 RA WISNIEWSKY C., BERCOVICI J., HENGSTENBERG C., AL-MHADAMI S.;
 RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and
 RT identification of a hot spot for mutations in the beta-myosin heavy
 RT chain gene.";
 RL J. CLIN. INVEST. 92:2807-2813(1993).
 RN [17]
 RP VARIANT HCM TRP-403.
 RX MEDLINE; 94093568.
 RA MOOLMAN J.C., BRINK P.A., CORFIELD V.A.;
 RT "Identification of a new missense mutation at Arg403, a CpG mutation
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in
 RT hypertrophic cardiomyopathy.";
 RL HUM. MOL. GENET. 2:1731-1732(1993).
 RN [18]
 RP VARIANT HCM ASN-615.
 RX MEDLINE; 93038688.
 RA NISHI H., KIMURA A., HARADA H., TOSHIMA H., SASAZUKI T.;
 RT "Novel missense mutation in cardiac beta myosin heavy chain gene
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 188:379-387(1992).
 RN [19]
 RP VARIANT HCM GLY-778.
 RX MEDLINE; 93343938.
 RA HARADA H., KIMURA A., NISHI H., SASAZUKI T., TOSHIMA H.;
 RT "A missense mutation of cardiac beta-myosin heavy chain gene linked
 RT to familial hypertrophic cardiomyopathy in affected Japanese
 RT families.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 194:791-798(1993).
 RN [20]
 RP VARIANT HCM VAL-908.
 RX MEDLINE; 93168485.
 RA AL-MAHDAMI S., CHAMBERLAIN S., CLELAND J., NIOYANNOPOULOS P.,
 RA GILLIGAN D., FRENCH J., CHOUHDURY L., WILLIAMSON R., OAKLEY C.;
 RT "Identification of a mutation in the beta cardiac myosin heavy chain
 RT gene in a family with hypertrophic cardiomyopathy.";
 RL BR. HEART J. 69:136-141(1993).
 RN [21]
 RP VARIANT HCM TRP-719.
 RX MEDLINE; 95179132.
 RA GREVE G., BACHINSKI L., FRIEDMAN D.L., CZERNIOWICZ G., ANAN R.,
 RA TOWBIN J., SEIDMAN C.E., ROBERTS R.;
 RT "Isolation of a de novo mutant myocardial beta MHC protein in a
 RT pedigree with hypertrophic cardiomyopathy.";
 RL HUM. MOL. GENET. 3:2073-2075(1994).
 RN [22]
 RP VARIANTS HCM CYS-513; ARG-716, AND TRP-719.
 RX MEDLINE; 94110336.
 RA ANAN R., GREVE G., THIERFELDER L., WATKINS H., MCKENNA W., SOLOMON S.,
 RA VECCHIO C., SHONO H., NAKAO S., TANAKA H., MARES A. JR., TOWBIN J.A.,
 RA SPIRITO P., ROBERTS R., SEIDMAN J.G., SEIDMAN C.E.;
 RT "Prognostic implications of novel beta cardiac myosin heavy chain gene
 RT mutations that cause familial hypertrophic cardiomyopathy.";
 RL J. CLIN. INVEST. 93:280-285(1994).
 RN [23]
 RP VARIANT HCM THR-797.
 RX MEDLINE; 96047155.
 RA MOOLMAN J.C., BRINK P.A., CORFIELD V.A.;
 RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
 RT myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL HUM. MOL. GENET. 6:197-198(1995).

[24]
 RN VARIANT HCM CYS-453.
 RP MEDLINE; 96209901.
 RX RA KUN P., WU C.-W., LIEN W.-P., LIEN C.-C.,
 RA KUO Y.-L., CHEN J.-J., TANG T.-K., CHENG J.-J., LIN S.-Y., LIOU Y.-C.,
 RT "Malignant familial hypertrophic cardiomyopathy in a family with a
 RT 433A>G mutation in the beta-myosin heavy chain gene:
 RT coexistence of sudden death and end-stage heart failure."
 RL HUM. GENET. 97:585-590(1996).
 RN [25]
 RP VARIANTS HCM THR-349 AND TRP-719.
 RX MEDLINE; 98204402.
 RA JESCHKE B., UHL K., WEIST B., SCHRODER D., METTINGER T.,
 RA DOHLEMANN C., VOSBERG H.-P.,
 RT "A high risk phenotype of hypertrophic cardiomyopathy associated with
 RT a compound genotype of two mutated beta-myosin heavy chain genes."

Query Match 1.9%: Score 146; DB 1; Length 1935;
 Best Local Similarity 16.9%; Pred. No. 1.2;
 Matches 238; Conservative 207; Mismatches 529; Indels 434; Gaps 51;

QY 203 DYGEPPVDRNAVVO-----NFMKDSGKWTTSGLNFTDPADQQLHEF 248
 DB 461 DLAGEIENFENFQDLCINTEKLEQEFNNHMFVLEQEEYKEGIEWTFIDFGMDLQAC 520
 QY 249 HMDNDKLVG-----NLTWRIKTLDDGFTFRSAVOTFVQDPLPGALYHND 293
 DB 521 IDLIRPMGIMSLIEECMFPRATDMTERKAKLEF-----NHL 557
 QY 294 NEFLHODOWYILTSQIEKRPD-DYIFVYGRND-----AMDYGSVYITRSPILPESI 348
 DB 558 GKSANFQK-----PRNIKKRPAHSLIHVAGIVYNIIGW-----LQKKDPLNETHY 605
 QY 349 IPNLQKAA-KSYGRDFNNEITDNGSGPEPLVERLEKTAEGEELLKEAVEIEVEK 407
 DB 606 VGLVGSLSKLSTLFANY-----AGADAP-IEKGKGAKKSSFOYSAH-RENNLK 657
 QY 408 EYEVKADTENTLFQRLBEFKELODEBNFVRELSKEKE-----LNELEMEAT- 458
 DB 658 LMTNLRST-----HPHFVRCILIPNETKSPGVNDNPLVMQKCNVYL 699
 QY 458 EYEVKGRALP-----IRKLMAAPHSNFIANHETIKYVSGKLPGHRSWG 506
 DB 700 EGRICRKGFPNNILXGDRQRRILNPAAPGQFIDSRKKAEXKLSLIDHNOYKFG 759
 QY 507 WEDYFGSIIVAKICSSRRIPRYFRKSPRICCGIDSGLOLFSHGHNLSPAS---INON 563
 DB 760 HTKVFYFKAGLIGLEEMRDELSRIITRI--QAOSRGVLAEMEYKKILERRDSLAVTQWN 817
 QY 564 VPKNGSGCKFPKVALMWEKMGOFKATAIVAFILISVASKADAVDLTCTCLKEC-- 622
 DB 818 IR-----AFMGVKNM--PMKMLYFKIKPILKSKEREKMSKMEFFRLKEALE 864
 QY 622 -----RLEIAKICSNPACANVACLOTNNRPDETECOIKCGDLFENSV-----VDEINE 671
 DB 865 KSFARKELEKKNVSLLOEKENDILOQAOQMDLADAEKCDQLINKKIQLKAEKEMNE 924
 QY 672 -----CAVSRK--KCVPRKSDYGFPPVDPSTVLVOKFMKDSFGKWFITRGL 716
 DB 925 RLEDEEMNAELAKRKLDECESELRDIDDELTLAKYEKK----- 969
 QY 717 NPFEDADQCLHEFHTEENKLVGLSWIRTPPGGFFTRSAVOKFVODPKYPIILYHNH 776
 DB 969 -----HATENR-VKNLTPEMAGLD-----ELIAKLKREKALOEAOHQAOL 1007
 QY 777 EYLLYODWYILSKVENSEPDIYFYVYKGRNDAMGYGGSVLYTRSA----- 825
 DB 1008 DDLQAEEDKVTILTKAKVYLEQV-----DDELSLEQEKQVVMDELRANKRLE 1056
 QY 825 ---VLPESIPELOTAQKYG-----RDF-----NTFIKIDNCGPE-----PPLVE 863
 DB 1057 GDLKTGESIMDLNDKQOOLDERLKKRDFELNALNARIEBOALGSOLOKKLELQARIE 1116

QY 864 RLEKVEEGERT-----IIKEVEEIEEVEK----- 890
 DB 1117 ELEFEL-ESEERARAKVEXLRSDLSRELEISERLEFAGATSVQIENKKREAFQMR 1175
 QY 890 -----VRDEKVTLSKLEFEGKELQORDEENFLRLSEEDVDGLKME 933
 DB 1176 RDLLEATLOHEATTAALAKRKHADSVALEQIDNLORVOKULEKESSEFKLELD---V 1231
 QY 934 ATVEVLFGRALFIRKLMAVATHCETSPCHDIRFFSSDDGIGRGITRRKINGTFLLKI 993
 DB 1232 TSMNEQITAKANLEKM-----CRT--LEDOMNEHS-----KAEFIQORSVNDL- 1274
 QY 994 LPEIQGADLRTYSGRSRRLP---SAFRSGFSKGIPIYPLPSKNELKELTAPDLIKLVG 1050
 DB 1274 --TSQAKAJOIENGELSRQDLDEKALISQUTRG-----KLITYOQLEDLKRQLEEVAK 1326
 QY 1051 LACAFILVSAVDALAKTACILKACRIELAKCIANPACANVACLOTNNRPDETECO 1110
 DB 1327 NALHALOSARHDCDLRLQYEETETAKAELQRVLSK--ANSEVIAQWR----- 1374
 QY 1111 IKGGDLFENSVDENFECASVRKKCVPRKSDIGFPPADPSVLVQNFISDENGKWTYS 1170
 DB 1374 -----KLETDG:QRTLELEAKKLAORLOPAE-----AVE 1405
 QY 1171 GLNTPEDAPDQQLHEFHTEGDNKLAVGNISWRIKTLDGFFTRSAVOKFVODPNQGVLYN 1230
 DB 1406 AVAKKCSLEKTKHRLQNEIEDLMVD-----VERSNAAAALDKKO---RN 1448
 QY 1231 HDNEYLHODOWYILTSKIKNEPDIYFYVYGRNDAMGYGGAAYVYTVSSVPLNIIPE 1290
 DB 1449 FDKILAEKMRKQRYEESQELSSQKE-----ARSISTELFLKNAYES 1491
 QY 1291 LE-----KAKSIGRDFSTFIRDTMTCGPALVERIEKTYEEGERIYVEVEEIEEVE 1344
 DB 1492 LEHLETKRENKNIQIEISLTLEQOLSSGKTHLEKVRQLOAEKMEQSALEBAEASL 1551
 QY 1345 EKEVEKGTENTLFO-----RLAGFNELEKODEBNFYRELS----- 1382
 DB 1552 EHEGKILRAQLFENQ:KAEIERKLAEKDEMEQAKRNHLRVVDSLOTSIADETRGRNEA 1611
 QY 1382 ---KEEME-FLDEIKMEASEVEKLFQKA 1405
 DB 1612 LRVKRMKMGDLNEMEIQLSHANRAEA 1639

RESULT 6
 ID NUM1_YEAST STANDARD: PRT: 2748 AA.
 AC Q00402;
 DT 01-DEC-1992 (REL. 24, CREATED)
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE NUCLEAR MIGRATION PROTEIN NUM1.
 GN NUM1 OR YDR150W.
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCCHAROMYCETALES;
 CC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 28362 / FL100;
 RX MEDLINE; 92079907.
 RA KORNANEC J., SCHAAF-GERSTENSCHLAGER I., ZIMMERMAN F.K.,
 RA PERECKO D., KUENTZEL H.;
 RT "Nuclear migration in *Saccharomyces cerevisiae* is controlled by the
 RT highly repetitive 313 kDa NUM1 protein."
 RL MOL. GEN. GENET. 230:277-287(1991).
 CC -I- FUNCTION: CONTROLS NUCLEAR MIGRATION. NUM1 SPECIFICALLY CONTROLS
 CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
 CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
 CC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
 CC ENVELOPE.
 CC -I- ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT CONSENSUS
 CC (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH FLANKING

CC DOMAINS OF THE TANDEM REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X61236; G4072; .
 DR PIR: S19052; S19052.
 DR SGD: L0001287; NMU1.
 DR PROSITE: P55003; PH-DOMAIN: 1.
 DR PFAM: PF00169; PH: 1.
 KW REPEAT.
 FT DOMAIN 593 1384 12.5 X TANDEM REPEATS.
 FT REPEAT 593 656 1.
 FT REPEAT 657 727 2.
 FT REPEAT 728 798 3.
 FT REPEAT 799 862 4.
 FT REPEAT 863 926 5.
 FT REPEAT 927 990 6.
 FT REPEAT 991 1054 7.
 FT REPEAT 1055 1118 8.
 FT REPEAT 1119 1182 9.
 FT REPEAT 1183 1246 10.
 FT REPEAT 1247 1310 11.
 FT REPEAT 1311 1374 12.
 FT REPEAT 1375 1384 13 (INCOMPLETE).
 FT DOMAIN 2573 2683 PH.
 SQ SEQUENCE 2748 AA; 313202 MW; 3085662C CRC32;

Query Match 1.9%; Score 143; DB 1; Length 2748;
 Best Local Similarity 17.9%; Pred. No. 2.9;

Matches 283; Conservative 231; Mismatches 582; Indels 486; Gaps 72;

QY 62 SFSQSSSHCKKCKSQICSDTSEFELQRPDLKRGMTLLEKQMOF-----10LAIVLC 115
 Db 748 SAYEDLVCKCKENPDV-----EFLKESAKLGHITVSSSEYSELQRYSELEKEVEQS 800
 QY 116 TFIIVPRDAVDALCTACILKECRIFELAKIANPSC-----AANVACLOTNNPDET 169
 Db 801 LAYIVHAKATDH-----HLSDSAYEEELVCKENPDMEFLKESAKLGHITVSSNEAYSEL 856
 QY 170 ECQI-----KCGD--LFENGSVVDQFNECAVSRKKCVPRKSDVGEFVPDRNAV 215
 Db 857 EKKLEOPSLAYIVHAKATDHLLSDSAYEDLYKC-----KENS DV-EF----- 900
 QY 216 VQNNMKMDFSGKWTITSGLNPTFAPDCQIHEFHMKDLVGNITWIKITLDGFFTRSA 275
 Db 900 -----LKEKSAKLGHTVSSNEAYSELEKLEQ-----PSLAYIVHAKATDHLLSDSA 948
 QY 276 VQTFVDPDIPGALYNHNEFLHY-----DWMYILSSQIENKPPDYIFVYTRG 324
 Db 949 YE-----DLVCKENPDMEFLKESAKLGHITVSSNEAYSELEKLEQPSLEYVHAKA 1002
 QY 325 RND--AMDGYSV-----LYTRSPITPESITP-----LQKAKSVGRF--- 364
 Db 1003 TNHLLSDSAYEDLVCKENPDMEFLKESAKLGHITVSSNEAYSELEKLEQPSLEYV 1062
 QY 364 -----NNFITDMS-----CGEPPLVERLEKTAEGEKLITKEAV-EIEBEVKE-- 409
 Db 1063 HAKATNHLLSDSAYEDLVCKENPDMEFLKESAKLGHITVSSNEAYSELEKLEQPSLE 1122
 QY 409 --VEKVDTEMTLFORLEGFELQDDEFNFVRLSKERKILNE--LQMEA-TEVEKLF 463
 Db 1123 YIVHAKATNHLLSD--SAYEELVCKENPDVEFLKESAKLGHITVSSNEAYSELEK-- 1179
 QY 464 GRALPIRKLKLMALAPHSNPLANHETIKYVYSKLPGRKRFSGWEDYGSIVAVK----- 519

Db 1179 -----KLEQ-----PSLAYIVH-----AKATDHLLS-----DSAYEDLVCKENPDV 1217
 QY 519 -----ICSSRIPIRYFRKSPRIOCGLDSGL-QLSHGKHNLSPHASINQV 564
 Db 1218 EFLKESAKLGHITVSSNEAYSELEK-----LEOPSLAYIVHAK--ATDHLLSDSA 1268
 QY 565 PRKNSGCKFPDVALMWEKMGQFAKTAIVAFILSVASKAD-----AVDAKTCCTCL 617
 Db 1269 YEDLVCKENPDMEFLK-EKSAKLGHTVSSNEAYSELEKLEQPSLEYVHAKATNHLL 1327
 QY 618 LKECKLE-LACISNPAC-----AANVACLOTNNPDETECOIKGDLFENSVYDEN 670
 Db 1328 LSDSAYEDLVCKENPDMEFLKESAKLGHITVSSNEAYSELEKLEQPSLE----- 1378
 QY 671 ECVAVRKKCVPRK-SDVGF-----VPPPSVLQKFKMDKSSGKWFITRGLNPTFPAF 723
 Db 1378 EYLVAHQAQIOSKLIISIDFNTLANPSMEDMASKLOKLEQIVSNDXYI--ALKNTMEKP 1435
 QY 724 DCQ-----LHEFH-----TEENKLVGNLSWRIRTPDGGFFTRSAVQKFOVD-PRYPGILYN 773
 Db 1436 DVELLRKLGKXHIIDTTYNELVSNFN-----SPTLKFIEKAKSGRYLI 1482
 QY 774 HDNEYL-----LYODWYILSK--VENSPE-----D 798
 Db 1483 EPNEYLDLNRATLTPPSKEEIDNEFCQIGYALDSKEVERLRKLSLENPSSKFFTEENALLD 1542
 QY 799 YIFV---YKGRNDAMDGYSGLVYTRSAVLPESITPELOTAQVGRDFNFITDNTC 855
 Db 1543 LVLYDKTEYOAKD-----NASNKSLLPSTKVL-----DEVFM----- 1577
 QY 856 GPEPPLVERLEKKAEGERTIIEVEEIEVEKVADEKVTLFKLFEGFKLQD----- 912
 Db 1577 -PAPQLASAKESLPQK-----RTLSIDINELKAL-----GYAIRKENLPN 1616
 QY 912 -EENFLELSEKEM-----DYDLGKMEATEVEKLEF-----RALPIR 948
 Db 1617 LEKTIYDMAKNDVLMCSKFSVLPLSTEEYDNNRKEHTKIINIGDPSIDFLKKECKEY 1676
 QY 949 KLMVAATH-----CFTSPCHDIRFSSDDGIGRLGTRKRINGTLLITPIIOSAD 1001
 Db 1677 QMLISKHDEKOEALENBQYFIEKASALGY-----ELVEVELDRKQKOWITSPD 1729
 QY 1002 LRTTGGSSPPLSAFRSGSGKGFIDVLPSSKNELKETAPLLKLVGLVACAFIYVSA 1061
 Db 1730 IDYQEKRAAN-----EMVLL--RNEKE--ALOKKIYPSLTILIEKAA 1770
 QY 1062 -----DAVDALCTACILKCGRIELAKCIANPACAAVACLOTNNRPDETECOIKG 1114
 Db 1771 GANKIILVDOIEYDET-----IRKC-----NHPTRMELESCH 1802
 QY 1115 LDFENGSVDFNECAVSRKKCVPRK-SDGEPAPDPVVLQNNPISFNCKWTITSGLN 1173
 Db 1803 HL--NVLDDQNFYSLTRLEPLKRNVEDLINTLSLNYIAIPNTIYODLCKYE--N 1855
 QY 1174 PTEDAFDCQLHEFHTEGDKLVGNISWRIKTLDGSEFFRSVAVQFVDDPNPGVLY-- 1230
 Db 1856 PNDYDKDSLK-----MDYAIISROYELAMVAKYKQPQDIYAKIS 1896
 QY 1230 -----NDNEYLHYQDDWYILSKIEKRPDYIFVYVYRGHNDAMD 1270
 Db 1897 SEKIDHIVPLSEYNLMWNTYRNPSLSYKAEKAVLNHILIKEDDYKNIL-----AVSE 1950
 QY 1271 YGGAIVYTRSSYLPSIIPLELEKAKSGRPSFIRINDTCGEPPLVERIEKTYEEGE 1330
 Db 1951 HPIVHILSRHLLKLVLD-----RNDFAVMSRSIE--KPTIDELSTALSMG- 1998
 QY 1331 RIIVKEV-----EETEEVEKEVEKVGRTMTLFORLAEGENLKQDEENVPRLSYEEM 1385
 Db 1998 KILVNSTHKNRKLSEDSDFLTKAKQDGLIISKEYISELRDQIDRPNLDVILKEKA 2057
 QY 1386 EPIDELTKMASVEKLFQKALP 1407

QY 663 NSV-----VDFEN-----CAVSRK---KCVPRKSDVGFPPVDPBSVLQKED 702
DB 911 NKIOLEAKVKEMTERLDEDEEMNAELTAKRRKLEDECESELRDIDLELTLAKVEREK-- 969
QY 703 MKDESGKWFITRGNTFDFADQCLHFEFTEENKIVGNLSMRTRTDGGFFIRSAVQKRV 762
DB 969 -----HATEK-KVNLTEENAGLD-----EITVKLT 993
QY 763 QDPYFGILVHNDVYLLQDDWYIISKVENSPEYIFVYKGRNDAMDGGGYLYTR 822
DB 994 KEKRALOEHOALDLDQAEEDKVNILTRAKVKLEQOV-----DLEGLSDQK 1042
QY 823 SA-----VLPESTIPELQTAQVQ-----RDF-----NTFKTNDTCGP 857
DB 1043 KVRMDLERAKRKLEGLDKLTQESTIMDENDKQDLERLKKDFELMALVARIEDQALGV 1102
QY 858 E-----PPLVERLEKKEVEGERT-----IKEYEELIEEVEK----- 890
DB 1103 QLOKRLKELQARIELEEL-FAERIRAKAVEKLRSDLSRELEISERLEEGATVQI 1161
QY 890 -----VRDEVTLSKLPESGFKELQDDEENFLREL 919
DB 1162 EMNKKRAEFOKMRDLLEATLQHEATAALRKHADSVAELEQIDNQLRVKQKLEKEX 1221
QY 920 SKEEMVDLQGLKMEATEVEKLFGRALPIRKMAVATHCTSPCHDIRREFSSDDIGRLG 979
DB 1222 SEPRLELD-----VTSMEQIITAKANLEK-----CMT--LEDQNMHRS-----KAE 1264
QY 980 ITRKINGITFLKILPPIQASDIRTGGSSRPL---SAFRSGFSKGFIDVILPSPKNE 1036
DB 1265 ETORSVNDL-----TRPRAKLQTEGELSRLQDEKELISOLTRG-----KLITYQOL 1312
QY 1037 KETLAPILKLVGLACAFIIVSADAVATLCTOCLKGCIEELAKCIANPACANVAC 1096
DB 1313 EDLKRQLEEBVAKKNAHALASARHCDLLQEQIEETEAALQRYLSK--ANSEVAQ 1370
QY 1097 LQTCNNRPDETECOIKCGDLFENSVDDEFENEGAVSRKCKVPRKSDLGEPAPDSVLQV 1156
DB 1371 WRT-----KYETDAIOTPEELEAKKLAQLODAE----- 1403
QY 1157 FNISDFNGKMYIISGLNPTFDADFQOLHFEFHESGDKLVGNISMRKILDSGFTTSAAVQ 1216
DB 1403 -----AVEAVNAKCSLEKTRKRLQNEILEDLVND-----VERSNAA 1438
QY 1217 KEYQDPNQPQVLYNHDNEXLHAYQDDWYIISKIENKPEYIFVYKGRNDAMDGGYGV 1276
DB 1439 AALDKQO-----RNFDKILVEWKQKEYESQSELSQKE-----ARSL 1477
QY 1277 YRRSSVLPNSIPELE-----KAAKSTGRDPSTIRIDNTCGPPALVERIEKTVEEGE 1330
DB 1478 STELTKLNAVEESLEHETFKRENKNLOEISDLTEOLGSTGSIHELKIRKQBLAEK 1537
QY 1331 RIYVEVEEIEEVEKEVEKVGRTMTLFO-----FLAGFMLEKODENFVREL-- 1382
DB 1538 LELOSALBEAASLEHEESKILIRAOLEFNQIKAELEKRLAKDEMDQAKNHLRVVDSL 1597
QY 1382 -----KEEME-FLDEIKEMASEVEKLFGRK 1405
DB 1598 QTSLDAETRSRNEALRVKMKMGDLEMEIOLSHANRMAEA 1639

CC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE: 93247549.
RA KOELING R., NGUYEN T., CHEN E.Y., BOTSCHEN D.;
RT "A new yeast gene with a myosin-like heptad repeat structure."
RL MOL. GEN. GENET. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 9420526.
RA BOU G., ESTEBAN P.F., BALADRON V., GONZALEZ G.A., CANTALEJO J.G.,
RA REMACHA M., JIMENEZ A., DEL REY F., BALLESTA J.P.G., REVUELTA J.L.,
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UB12 and MPl1 genes and three
RT new open reading frames."
RL YEAST 9:1349-1354(1993).
CC - FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC - SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC - CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L01992; G171959; -
DR EMBL: X73541; G450554; -
DR EMBL: Z28320; G486587; -
DR PIR: S38173; S38173.
DR SGD: L0001122; MPl1.
KW MYOSIN; HEPTAD REPEAT PATTERN; COILED COIL; DNA REPAIR.
FT DOMAIN 69 487
FT DOMAIN 531 1678
FT DOMAIN 1834 1866
FT COILED COIL (POTENTIAL).
FT COILED COIL (POTENTIAL).
FT CONFLICT 301 301
FT R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 8B01FD0D CRC32;

Query Match 1.88; Score 137.5; DB 1; Length 1875;
Best Local Similarity 17.58; Pred. No. 3.6;
Matches 252; Conservative 240; Mismatches 525; Indels 421; Gaps 67;

QY 171 COIKCGDLFENSVD-----OFNEC-----AVSRKCVPRKSDVGEPP 208
DB 30 CSLBQVKSFDGDVYKHLNDKILQFNLKSENKLVYSPDELKASSLTKKIDGLKTEW-ENW 88
QY 209 VPDRAV-----VQNFNMKDESGKWTGSGNP-----TFDAF 241
DB 89 IRENDKIRKERNDITFVKESEVENKMKLSSELEFVARRKLDLLEKKETQSNQOQTLKIL 148
QY 242 DCOLHEFH-----END-----KLVGNLTWRKTLDSGFT-----RSAQTVQDPDLG 287
DB 149 DERLKEILEIVAVENNRNSCKKLRSTIMDLTKQOGYITNDLNSRTLEERTQDLTL-- 207
QY 288 ALVHNDNEFLHAYQDDWYIISQIENKPPDYIFVYKGRNDAMDGGGYIYRSPTLPES 347
DB 207 -----LOSNDW--LEKELRSKNEQYL--SYRQKTD-----K 234
QY 348 IIPNLOKAASVGRDF-----NMFITDSCGPEPPLVERLEKTAEBEGKLLIKAVIE 402
DB 235 VIIDIRNELNRLRNDPOMERTNDVLRQKN-----NELSKSQ--EKLQ--EIKELS 282
QY 403 EEEVEKAVRTDMLTFLOR-----LLEGFKELQODEENFVREL-----SKREK 446
DB 283 DSLNSEKQEF-SAEMSLKQRLVDLLESQLANVAKEELNSIRELNTAKVLADSKKOTPENE 341
QY 447 EILNELQMETVEKELFGRLPIRKLR-MALAPHSNLANHETIYVYGSKLPGHKRFSW 505

```

Db 342 DLKELQITREKLAQCEKECELRISITDEADENENLSAKSSSDIFLKKQILKERITKE 401
QY 506 GWEDYFSGIVANICSSRRIPRYFRKSPRICQGLDSRGLOLFSGHKHNSPAHSINQNP 565
Db 402 HLGNOIETFEIIEHKKVPIINSFKERTDMLNNAALLHETSSEKNAKVELN---A 458
QY 566 KQNSGCKFPDVALMWEEKV-----GQFAKATAYAFILSTVASKA 605
Db 459 KNOKLVECENDLQITKQRLDLCRQIOYLLITNSVSDSKGPIRKEBIO--FIQNMQED 516
QY 606 DAV----DALKTCTCLKECR--LELAKCISNPACANVACLOTNNRPDETCQIKG- 659
Db 517 DSITTEDSOKVYTERLVEKKNIIQLOE-----KNALLVVRNLADKLESKEKSK 568
QY 659 ---DLFENSVDDEFNECAKCKVPRKSVGD-----FPVDPESVLQKED 702
Db 569 QSLQKIESEFVNEAKEAIIITLKS---EKMDLESRIEQLQELKTSVNEBASVSVNT 625
QY 703 MKDSGKMFITRGINPFDFAFDQHEFHHEENKLVGNLSMRIRTPDGGFTRSAQKVEY 762
Db 626 IKQ-----LLETKRDLSEVOLOTRISQIT-----RESTEN-----MSLNKEI 665
QY 763 QDPKPYGLYNHNEY-----LLYQDDWYLLS-----SKVENSPEYIIVYK 806
Db 666 QD-----LDSKSDISIKLKEKSSRIABERFKLSNLTDLTKAENDQLRKFDYLQ 719
QY 807 -----RNDAMDGYG-----SVLYTRSAVLEPESITPELOTAQKVGDFNFIETD- 853
Db 720 TILKQDSKTHETLNEVYSCSKSLIVETELNLKEE-----QKLVHLEKNQEL 770
QY 853 NTGCGE-----PLVERLEKVEEGEPTIIEVEEIEEVEKVRDKEVTLFSKL 901
Db 771 NKLSPKSLAIWYLOLOTLQKEREDELLETRKSCQKKIDLELALBELK-KER---SQK 826
QY 902 FEQGEKLORDE-----ENFLRELSKEEMVDLGLKMEATEVEKLFGRALPIRLMAVA 954
Db 827 DHHQLLEEDNNSNIEMVQNKIEALKDYESVITSVKQDIKLOKRYKASLEKEIE-- 885
QY 955 THCFSPCHDRIRFES-----SDQIGRLGTRKIRNGTFLFKILPIQSDALRTG 1006
Db 885 -----EDKIRLHTYVNMDETINDSL--RKELEKSKINLTDAVSQIK--FYMDLLET- 933
QY 1007 GSSRPLSAFRSGFGKIFDIIVPLPSKNELEKELTAPLLKLVLGVACAFILVPSADAVDA 1066
Db 933 ---TSQSLOQTNSKIDESFKDFT---NQINKLT-----DEKTS 964
QY 1067 LKTCACLLKGRIELAKCIANPACANVACLOTNNRPDETCQIKGDLFEN-----SV 1121
Db 965 LEDKISLLEKQMFNL-----NNELDLQKKMEKEREKADFKKRISITLQNNKREVAV 1014
QY 1122 VDEFNDCAVSRKKCVPRKSDLGEPAPPSVLYVONFNISDFNGKYIITSGNPTFDADC 1181
Db 1015 KSEY-----ESKLSKIQMDL-----DOQIYANTAOONNYOEOLQKADVSKITSELR 1062
QY 1182 QLHEFHTEGDNKLVGNISWRIKTLDGSEFTRSAYOKVQDPQPVLYNHDEYLAHODD 1241
Db 1063 QLHTY-----KQGVKTLN---LSRQLENALKENKS---WSSQKESLLEQID 1104
QY 1242 WYLLSKIKNKEP-----DYFFVYIRGRNDAMDGYG-----AVVYTRSSVLPNSIIP 1289
Db 1105 ---LSNSRIEDLSQNKLYDQIOIYTAADKEVNNSNNGGLNNIITLRERDILDTFYT 1162
QY 1290 ELEKAKSKIGRDESTF-----IRTDNCGEPALVRIEIKTEVEEGRITVKEVEEIE 1341
Db 1163 VAEKDKMKLRQKISLMDVELQDARTKLDNS-----RVKQ---ENNSSIIQHQDDIM 1210
QY 1342 EYEVEKEVEKVGRTENTLQFORLAEGFENELKODEEVEVRELKSEMEFLDEIKMEASEVE 1399
Db 1211 EKL-NQNLILRESNITL-----RNEL-ENNKKRKELOSE---LDLKKONVAPTE 1255

```

RESULT 9
DMD_CHICK

```

ID DMD_CHICK STANDARD; PRT: 3660 AA.
AC P11533:
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE DYSTROPHIN.
OS DMD.
OS GALLUS GALUS (CHICKEN).
OC EURAYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAL; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89098331.
RA LEMAIRÉ C., HEILIG R., MANDEL J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA.";
RN NUCLEIC ACIDS RES. 16:11815-11815(1988).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE-MUSCLE.
RX MEDLINE: 89210800.
RA LEMAIRÉ C., HEILIG R., MANDEL J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal coding and 3' untranslated regions between man and chicken.";
RN EMBO J. 7:4157-4162(1988).
CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE PLASMA MEMBRANE.
CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN, ABB-120, ABB-180, OR BETA-FODRIN).
CC -1- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC CC
CC EMBL: X13369; G63370; -.
CC PIR: S02041; S02041.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS00020; WW_DOMAIN_2; 1.
DR PRAM: PF00307; actinin-binding; 1.
DR PRAM: PF00397; ww_rsp5_wmp; 1.
DR PRAM: PF00435; spectrin; 22.
DR PRAM: PF00569; 22; 1.
DR HSP: 001082; IAA2.
DR STRUCTURAL PROTEIN; ACTIN-BINDING; CALCIUM-BINDING; CYTOSKELETON; KW REPEAT.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 300 3000 26 SPECTRIN-LIKE REPEATS.
FT DOMAIN 3052 3085 WW DOMAIN.
FT DOMAIN 3086 3357 CYS-RICH.
FT VARIANT 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; AF61A205 CRC32;

Query Match 1.8%; Score 135; DB 1; Length 3660;
Best Local Similarity 17.6%; Pred. No. 12;
Matches 245; Conservative 196; Mismatches 482; Indels 472; Gaps 64;

QY 71 KDKSOISQISDFEIORFPLKRGM--TLF-LKQWROF-----IQLATIVLC 115
Db 1923 KKKEDLNANVQEARLKDGAQKAVEPTVLQSKRMRDFESKRAQRRNLNVAIQIVLBD 1982
QY 116 TTVIVPRVDAVDALKTCACLLKECRILACIANPSCANVACLOTNNRPDETCQIK 175
Db 1983 TTFVWTESMVETVYVSTYLAET-LQLLQALSEVERLNSPYLOAKD-----C 2030

```

```

QY 176 GDLFENSVDQFNECAVSRKKCVPR-----KSDGEEFVPPRANAVONFNKNDF 224
  |||
Db 2031 EDLLQO-----EELKNIKIDGLRLOGHIDI IHSKRTPALQSNTPRETANIOD-KILQL 2083
QY 225 SGRWYITSGL-----NPTD-----AFDCOLHEFH-----MENKL-----VGNLT 260
  |||
Db 2084 NSQMEKVMKMYRDQARDKSKEMKRLFHCEKMSFNEMLETEKLSRAQIEAGDVGHVK 2143
QY 261 WR--IKTLDGFFTSANQVTFVQDDLPG----- 288
  |||
Db 2144 TKQFLOELQDGI-----GRQQTAVKTLNVGEBIIEQSSAADANVLKEQLGNLNTRMQEIQR 2200
QY 288 -----ALYNHNEFLHYQDDWYILSSQJENKRPDIIFVYRGRNRMNGYCGSVIYTR 340
  |||
Db 2201 QVEKRRRIIEBKNTLSFQEDLNKLIMLEETENVIAIPLERQNEQDLRDCIGKVKLRV 2260
QY 341 SPTLESIIPLNQAASVGRDNFNFTTDSGCEPPELVERLEKTAEGEKLLEKEAVE 400
  |||
Db 2261 EELLPHKGI--LKRLNETG-----TTLGASLNERKHKLSTLKEASRLLVSRD 2311
QY 401 IEEVEKEVEKVRDTEMLTFORLLEKRELOQDEENFVRLSKEKEILNELOMATEVE 460
  |||
Db 2312 LPEK-QKEIF-----ILKDFLELNOQ-----INOLITLWITPVK 2344
QY 461 --KLFGK-----ALPIRLKLMALAPHSNFLANHETIKYVGSKLPGHKRFSGMWEDYFG 512
  |||
Db 2345 NQLEIYNQVQPGAFDIKETEA-----VQAKOPNVE----- 2378
QY 513 SIYAKICSSRIIRIFRKSRIQCCIGDSGLQGLFSHGKHNLSPAHSINQNPKNSGCK 572
  |||
Db 2378 --VLSKG-----HLTKERP-----ATHPVKKKLEEDLNADWK 2407
QY 573 FPKDVALMWEKMGQFAKIAIVAFITSVASKADAVDALKTCTCLKEGLRLACISNP 632
  |||
Db 2408 AINHLIILQIKER-PTGEPRALTSBGVLT-SGVIVAADT-----QARVKEITTSFTP 2456
QY 633 ACAANVACLOTCNNRPDETCQIKCGDLFENSVDDEFNEC-----AVSR--KKGVPKRS 684
  |||
Db 2457 EMPSV-----LLEVPALADFNKAMWELTDWLSRLDRLEIKARV 2495
QY 685 DVGDEPVDPSPVLQKEDKMGSGKFTIRGLNPTDADQCLHEFHHEENKLVNSMR 744
  |||
Db 2496 TVGDLIDINDMIITKQKANNQDLEOR-----RP-----OLDELITIAAQ-----NLNKK 2537
QY 745 IRPPDGGEFTFSRAVQK-VODPKYPGILYHNHNE-YLLQDDWYILSSVENSPEIDYFV 802
  |||
Db 2538 TSNQEARITITDRIETKIQSQWMDVHGYLQNRQOLHEMQKDTQMLEAKQF--AEQVLE 2594
QY 803 YKGRNDAMNGYGSVLYTRSAVLPEST-----IPELOTAQKVGGRDN- 847
  |||
Db 2595 QAKAKLESWK-----EISYVEALKKQNSKELKQFSKEIRQWQMNIGVNDVALKPYRDSA 2650
QY 847 -----TIFKTDNTCGPEPLVERLEKKEVEGEFTI-----IKVE 881
  |||
Db 2651 DQTRKVELMTDN-----NATWATINKRVSEREALESALLMJOEYUDLEKFLAMLEAE 2706
QY 882 EIEEVEKVRDKEVTL-----FSKLEFGKELQD-----EEN--FLRELKE 922
  |||
Db 2707 TIANVLQDATHKEKTELEDPQWRELMKQWQDQAEIDAHTDIFHNLNDNGKILRLSEGS 2766
QY 923 EMVY-----LDGKIMATEVEKLFGRALPIRKLMAVAHCTSPCHDRIRFFSSDDGIGR 977
  |||
Db 2767 EDVAVLQRLDNNRFRWSLKR--KSLNIRSHLEAST-----DQMKR 2806
QY 978 LGITRRKINGTEFLK-----LIPPIQ-----SADIKRTGSSRSPIASF 1016
  |||
Db 2807 LHLISQELLAMQIKEDDELKQAPIGSDIPIYQKQNDVHRTFKRELKKEVINNALETV 2866
QY 1017 RSGFS-----KGIFFDIYPLP--SKNELKELTAPLLKLVGLACAF--LIVPSAD--A 1063
  |||
Db 2867 RLFLADQVEGLEKVPYPERDLSPERAOANTYKVLRRQADQVREMDKLNLRSDMQKKI 2926

```

```

QY 1064 VDALKTCACLLKGCRIELACIANPACAANVACLOTCNNRPDETCQIK----- 1113
  |||
Db 2927 DDALER-----LQSLQZAMDELDLKLQOAEAFKSGMOP 2959
QY 1113 CGDLFENSVDENECAVSRKKCVPR-----SDLGEPPADPSVLYONFNIS--DFN 1163
  |||
Db 2960 VGDLLIDLSLDHDEKYEKVAEAVPLKREKHQVNELAHFRAP-PIQLSPYTLSCLEDLN 3018
QY 1164 GRWYITSGLNPTEDAFDQCLHEFHTE-GDNKLVGNISWIKTLDGFFTSRAVQ--KTV 1219
  |||
Db 3019 TRWV--LQVALDERLRQLHEAHRDFGP-----SQHFLTTSVQGPWERA 3061
QY 1220 QDENQPSVLYNHDNE 1234
  |||
Db 3062 ISPKVYVYINHEQTQ.3076
  |||

RESULT 10
MISP_HUMAN
ID MISP_HUMAN STANDARD; PRT; 1937 AA.
AC P13535; Q14910;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE MYOSIN HEAVY CHAIN, PERINATAL SKELETAL MUSCLE.
GN MYH8.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE: 90323631.
RA KARSOCH-MIZRACHI I., FEGHALI R., SHOWS T.B. JR., LEINWAND L.A.;
RT "Generation of a full-length human perinatal myosin
  heavy-chain-encoding cDNA.";
RL GENE 89:289-294(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE: 95324556.
RA JULIAN E.H., KELLY A.M., POMPIDOU A.J., HOFFMAN R., SCHIAFFINO S.,
RT STEWMAN H.H., RUBINSTEIN N.A.;
RT "Characterization of a human perinatal myosin heavy-chain
  transcript.";
RL EUR. J. BIOCHEM. 230:1001-1006(1995).
RN [3]
RP SEQUENCE OF 502-1937 FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE: 90235862.
RA BOBER E., BUCHENGER-SEIDL A., BRAUN T., SINGH S., GOEDDE H.W.,
RT ARNOLD H.H.;
RT "Identification of three developmentally controlled isoforms of human
  myosin heavy chains.";
RL EUR. J. BIOCHEM. 189:55-65(1990).
RN [4]
RP SEQUENCE OF 860-1937 FROM N.A.
RX MEDLINE: 89234168.
RA FEGHALI R., LEINWAND L.A.;
RT "Molecular genetic characterization of a developmentally regulated
  human perinatal myosin heavy chain.";
RL J. CELL BIOL. 108:1791-1797(1989).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
  HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
  AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- DOMAIN: THE PDOLKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
  CYCLES OF A 23-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
  CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PFM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
  AKTYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM)

```

CC AND 1 HEAVY MEROMIOSIN (HMM). IT CAN LATER BE SPLIT FURTHER INTO
 CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
 CC -1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
 CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE
 CC CONSERVED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M36769; G189034; -
 DR EMBL: Z38133; G558669; -
 DR EMBL: X51592; G29466; -
 DR EMBL: Y00821; G34864; -
 DR EMBL: M35250; G531200; -
 DR PIR: A30220; A30220.
 DR MIM: 160741; -
 DR PFAM: PF00063; myosin_head; 1.
 DR PFM: PF00612; IQ; 1.
 DR HSP: P13338; 2MYS.
 DR MOSIN: MUSCLE PROTEIN: COILED COIL; THICK FILAMENT; ACTIN-BINDING;
 KW ATP-BINDING; METHYLATION; ALKYLATION; HEPTAD REPEAT PATTERN;
 KW MULTIGENE FAMILY.
 FT DOMAIN 1 841 GLOBULAR HEAD (S1).
 FT DOMAIN 842 1937 RODLIKE TAIL (S2 AND LMM DOMAINS).
 FT NP_BIND 842 1937 COILED COIL (POTENTIAL).
 FT DOMAIN 181 188 ATP.
 FT DOMAIN 658 680 ACTIN-BINDING.
 FT DOMAIN 760 774 ACTIN-BINDING.
 FT MOD_RES 132 132 METHYLATION (H1-) (POTENTIAL).
 FT MOD_RES 698 698 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 708 708 ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 15 15 A -> R (IN REF. 2).
 FT CONFLICT 970 970 E -> Q (IN REF. 1 AND 4).
 FT CONFLICT 1072 1072 M -> N (IN REF. 3).
 FT CONFLICT 1247 1247 M -> H (IN REF. 1 AND 4).
 FT CONFLICT 1251 1251 MC -> DG6 (IN REF. 3).
 FT CONFLICT 1261 1261 E -> G (IN REF. 1 AND 4).
 FT CONFLICT 1297 1297 K -> Q (IN REF. 1 AND 4).
 FT CONFLICT 1378 1378 KY -> NT (IN REF. 3).
 FT CONFLICT 1505 1505 EN -> AH (IN REF. 1 AND 4).
 FT CONFLICT 1847 1847 E -> D (IN REF. 1 AND 4).
 FT CONFLICT 1914 1914 D -> H (IN REF. 1 AND 4).
 SQ SEQUENCE 1937 AA; 222762 MW; BDE9EC93 CRC32;

Query Match 1.8%; Score 134.5; DB 1; Length 1937;
 Best Local Similarity 17.0%; Pred. No. 5.5; 597; Indels 529; Gaps 67;
 Matches 280; Conservativity 243; Mismatches

43 IRSNN-----GYFSFLFTSYKTSFSDSHCKDQSCISIDTSFEIQRFDLKRGMTLI 98
 Db VVNDNSRPFKFIIRIHGIGTKLASADIEYLLKRSV-----TFQK----- 282
 QY LEKOWROFIOLA-----IVLVCT-----FYIVPRV-----AVDALKTC 132
 Db 282 AERSYHFFYQITSNKKRDLIELMLITNPPDYAFVSGEITVPSIDOEELMTDSIDI 341
 QY 133 ACLKEIRIELAKCIANPSCANAVACLOTNN--RPDETCQKCGDLFENSVDQFNE 189
 Db 342 LGTPEEKVSIYKLTGAVMHGNNKFKQKQREBAEPDGETVADKA-----AVLSLNS 395
 QY 190 CAVSRKCKVPRKSDVGEFPVADNAVQNNMKDSGK-----WYITSGLNPT----- 238
 Db 396 ADLKALCYPRVK-VGNEYVTKGQTVQOVYNAVAGALAKAYEKKMLMVTR-INQOLDIK 453
 QY 238 -----FDAFDOLHE---FHMENDLVGNLWIRITLILGSGFTSRAVO--TF 279
 Db 454 QPROYFIGVLDIAGEFIFDENSELOLCINFTEKLOQPFNNHMFVLEOEYKKKGKGIWTF 513

QY 280 VQ--DPDL-----PGALYN-----HNEFLHYODDWYILISQIENKPPDYI 318
 Db 514 IDEGMDLACILIEKPGIGFIELEECMFPAKATPTSRKNKYDHLKSNFQPK--- 571
 QY 319 FVYRGRNDAMDG---YGGSVIYT-----RSPLPESIIIPNLOKA-KSYGRFNNFI 367
 Db 571 --VVKGAHAEHSLIHAYAGTVYDYNITGWLDKNRDEPLNTVGLYOKSAMKTLASLFSTYA 628
 QY 368 TIDNSGCPPLVERLEKKAEGEKILLIENAVIEEVEKEVEKVRDTEMLTFOLELGEF 427
 Db 629 SAE-----ADSSAKKGAK-----KKGSSQIYSALFRENLKLM 662
 QY 428 KELODEENFREL-----SKEKEILNLOMEAT-EVEKLGRLAPR----- 471
 Db 663 TNRSTHPRHVRCTIIPNETKTPGAMHELVHLOKNGVLEGIRCKGFPRIILYDPK 722
 QY 471 ---KLMALAPHSNLANHETIKYVSKLPGRKRFSGMEDYF---GSIYVAKICSR 523
 Db 723 QRYKVLNASAIPEGOFIDSKASEKLASIDIDHTQYKFGHTKVFKAGLLGLLEMRDE 782
 QY 524 RIRYRFRKSPRIICGDSR-----GOLFSGKHNSPASHINQNP 565
 Db 783 KLAQITRTQAVRGELMRVEYQKMLQREALFCIOYNVRAFNKYH--WPMKLEFKIK 840
 QY 566 KNSGCKFPKDVADLWWEKWOFAKT-----AIVAI-----FILSVAS 603
 Db 841 PLKSAATEKEMNTMKE-----FOKTKDELAKSEAKRELEKRVTLLEKNDLOLOVOS 896
 QY 604 KADAV-DALKTCICLKECELEAKCISNACANAVACLOTNNRPDEECQKGGDLFE 662
 Db 897 EADSLADEERCDILKN-KIOLE-----AKIEVTERAEDEE----- 934
 QY 663 NSVVEFNECAVSRRK-----CVPRKSDVGEFPVPPSVLVQKDFDKDFSGKVFITRGLN 717
 Db 934 -----EINAEITAKKRLDECESEKIDDLDELLTAKVEK----- 972
 QY 718 PTFDAFDQLEHFEFTENKLVGNSWRIRPDGFGFTRSVQKFEVODPKYIGLIYHNE 777
 Db 972 -----HATENK-VKNLTEMAGLDETIAKLSKEKALOE-----THOOTID 1011
 QY 778 YLLVQDDWYILISSVENSPEDYIFVYKKGNDAMDGVGSGVLTBSAVLPESIIPELOFA 837
 Db 1012 DLOAEEDKVNITAKKRLQOV-----DLEGSLEOKKRLM-----DIERA 1054
 QY 838 AOKGRDFNFITKDNTCGPE---PPLVERLEK-----VE----- 871
 Db 1055 KRKLEGLK--IAOESTMDKMDKQOLDEKKEFEITSNLISKIEDEQAVEIOLQKIK 1112
 QY 871 -----EGERT-----IKEYEELIEVEYK----- 890
 Db 1113 ELQARIEELGEIEBAERASRAKAKORSRLSELEISERLSEAGATSQVLELNKREA 1172
 QY 890 -----VROKEVTLPSKLFEGFEKLOREENFLELSKEEMDVL 927
 Db 1173 EFQKLRDLEAVYLOHEAMVAALKKHAADMAELGEODINQKQKOLEKESLKNET- 1232
 QY 928 DGLKMEATEVEKLEGRALPIRKLANAVATHCFTSPCHDRIRFESSDDGIGRLGTRKING 987
 Db 1232 DDLSSNAEISAKGN--LEKM-----CR-----SLEDQVSEL--RTKEEQ 1269
 QY 988 TFLKILPPIQASDRLTGGSSRL--SAFRSGFSKIGIDIPLPSKNLKXLTAPPL 1044
 Db 1270 QRLINDL-TAQRARLOTAGEYSROLDKDLVLSQSS-----KASTOOIEELKQLE 1323
 QY 1045 LKLVGACAFILVPSADAVADLAKTCACILKGRITELAKCIANPACANAVACLOTNNRP 1104
 Db 1324 EETKAKNALAHLOSRHDCDLRLQYEEDEGAELOALRSK--ANSEVAQWMT----- 1377
 QY 1105 DETEQIKCGDLENSVVEFNECAVSRRKCVPRKSDIGEPAPDPSPVLVONFINSDENG 1164
 Db 1377 -----KYETDAIORTEELEEKAKKLAORLOEAE-----HVEAVNA 1412

QY 1165 KWTITSGNLTDPADFCOLHEPHTEGDNKLVGMSIRIKTLDSGFTTSRAVQKFDVDPNQ 1224
 Db 1413 K-----CASLETKRLOLNEVNDLMDVE-----RSNACALADKKO 1449
 QY 1225 PGVLVYHNDNYLHYODDWITLSSKIKENKPEDELYVYRGRNDAMDYGGAVVYTRSSVLP 1284
 Db 1450 -----RNFDKVLSMKOKYEFTQALBASQK-----SRSLSTELFKVKNYEEEL 1496
 QY 1285 NSIIELEKAKSIGRDFEIRFTDTCGEPALVARIKTYEEGRITIVKEVEETEEV 1344
 Db 1496 -DQLETLRNRKNLQOEISDLTEQIAEGKQJHELEKIKOVEOEKCEIOALAEAEASL 1554
 QY 1345 EKEVEKVGRTMTLFO-----RLAEGFNEIKODENFVRELS----- 1382
 Db 1555 EHEEGKILRIQLELNVKSEVDKIAEKDEEIDQLKRNHTRVVEIMOSTLDAEITSRNDA 1614
 QY 1382 ---KEEME-FLDEIKMAEAEVEKLFEGKAL 1406
 Db 1615 LRVKRRKESGLNEMETOLNHNRLAASL 1643
 RESULT 11
 MSP1_PLAFK STANDARD: PRT; 1630 AA.
 ID MSP1_PLAFK STANDARD: PRT; 1630 AA.
 AC P04932:
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
 DE (PMSA) (1190).
 GN MSP-1.
 OS PLASMODIUM FALCIPARUM (ISOLATE K1 / THAILAND).
 OC EUKARYOTA: ALVEOLATA: APICOMPLEXA: HAEMOSPORIDA: PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86136024.
 RA MACKAY M., GOMAN M., BONE N., HYDE J.E., SCAIFE J., CERTA U.,
 RA STUNNENBERG H., BUCARD H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 RT Plasmodium falciparum merozoites: studies at the genetic level.";
 RL EMBO J. 4:3823-3829(1985).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RA PAN W., TOLLE R., BUCARD H.;
 RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- PPM: MERZOITE SURFACE ANTIGEN CONTAIN THE SPOUNCE OF 83 KD, 42
 CC KD AND 19 KD ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X03371: G929798: -
 DR PIR: A25120: SAZOK1.
 DR PFM: PF00008: EGF. 1.
 KM MALARIA: MERZOITE: POLYPROTEIN: REPEAT: SIGNAL: GLYCOPROTEIN:
 KM TRANSMEMBRANE: GPI-ANCHOR.
 FT STGNL 1 19 POTENTIAL.
 FT CHAIN 20 1630 MERZOITE SURFACE PROTEIN 1.
 FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
 FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
 FT CARBOHYD 97 97 POTENTIAL.
 FT CARBOHYD 259 259 POTENTIAL.
 FT CARBOHYD 755 755 POTENTIAL.
 FT CARBOHYD 759 759 POTENTIAL.
 FT CARBOHYD 774 774 POTENTIAL.

FT CARBOHYD 835 .. 835 POTENTIAL.
 FT CARBOHYD 911 911 POTENTIAL.
 FT CARBOHYD 955 955 POTENTIAL.
 FT CARBOHYD 1049 1049 POTENTIAL.
 FT CARBOHYD 1156 1156 POTENTIAL.
 FT CARBOHYD 1165 1165 POTENTIAL.
 FT CARBOHYD 1436 1436 POTENTIAL.
 FT CARBOHYD 1517 1517 POTENTIAL.
 SQ SEQUENCE 1630 AA; 187289 MW; DD2F8628 CRC32.
 Query Match 1.8%; Score 134; DB 1; Length 1630;
 Best Local Similarity 18.0%; Pred. No. 4.7;
 Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70.
 QY 184 VDGFNECAVSRKCCVPRKSDVSGFFPYDPDNRNAVQNMADPFGKWTITSGNLTDPADFC 243
 Db 240 IENINELIESKKTIDK-----NNKATKEERKRLYQAYDLS-----IYNK 281
 QY 244 QLHEFHENDKLVNLTWRIKTLDGGFTRSAVQTE--VQDP-----DLPGALYNHNE 295
 Db 282 QLEEAH-----NISVLEKRIDLTKKNENIKELDKINELKNPPANGSGTPTLLDKKK 337
 QY 296 FLHYQDDWITLSSQIENKDDY-----IFVYRGRNDAMDYGGSVITYRSPILPESII 349
 Db 338-IEEHEKEIEIAKTIKENIDSLETPDLELEYLRKRNKIID--ISAKYETKESTEPNE-Y 394
 QY 350 PNLKAKASVGRDNFNFTDNSCGEPFLVRLKTAEEGSKLI-----KEAVEIEE 404
 Db 395 PNGVITPLSY-NDINNALWELNSFG--DLINFPYITPEPSKNITTDERRKFLNEIEK 450
 QY 405 VEKEVKVADTEKTLFQRLBSGFKLODEENFVLSKEKEILNEL-----QMEAT 457
 Db 451 IKIEKKRI-----ESDKKSYEDRSKSLNDITREYEKILNELIYDSKPNNDI 498
 QY 458 EYEKLFGR--ALPIRKLRLMALPHSNPLANHETIYYVGSLLPGIKRSGWEDY-FESI 514
 Db 499 NEFKMKGRKYSKYVEL-----THNITRASYENSHNH-EKLTALKT--MEDYSLRNI 549
 QY 515 VYAKICSSRIIRYRRKSRICGDSRGDLFSGKHNLSPAISINOVPRKSGC-- 572
 Db 550 VYEKEL-----KYKK-----NLISKTENEI--ELIVENIKKDEQLEF 586
 QY 572 -----KEPKDVALWMEKMGQFATAIYALFILSVASKAVALKTCICLKECRLE- 625
 Db 587 KITKDEKNPKDEILEV-----SDIYKVOQKVLW-----NKIDELKKTOLIKNVELKH 636
 QY 625 -----LAKCISN-----PACAA-----NYACIQTCNNRPDET 651
 Db 637 NIHVPNSYKQENKQEPYLLVLYLKKEDLKYFMKVESLNEEKNKIKTBSQSDNSEPST 696
 QY 652 EQQL-----KCGDLFENSVDDEFNECAVSRKCCVPRKSDVGDFPYDPDSVLVQ-- 700
 Db 697 ESELIGQATTGQOAGSALGEG--DSVOAQOPOKQAO--PVPVPEPAKQV 747
 QY 700 -----KFDKDSFGKWFITRGILNPTFDALFDQQLHFFHBEKIKVNLNLSWRI 745
 Db 748 TPPAVNNKKTENVSCLDYLE-----KLVEFL-----NYSIC 779
 QY 746 RTPDGGFTTSRAVQKFDVDPKPYGILYNHD--NEYLLODDWYILSSKVEN-----SPED 798
 Db 780 H-----KY--ILVSHSTNMKILKO--XKITKEBSKLSLSCDPLD 815
 QY 799 YIF-----YYKGRNDAMDYGGSV-----LYTSAVL-----PESIIPELOTAQ 839
 Db 816 LIFNIQNNIPIVY--SMFDSLNNSLSQLEMELEYEKEMVCNLYKLKDNKIKINLLEAK 871
 QY 840 KYGDENTFIKTDTGCEP-PLVERLEKRYEGERT-----IIRKEVEI----- 884
 Db 872 KV-----SISVKTLSSSMQLPSLIPQDKPEVSANDDTSHSTINLNSIKLFEINILSGKNK 927
 QY 884 -----EEVVKVADKEVTLFESKLFEGFKELQDEENFELRSLSEKMDVLDGLK 931

[illegible]

```

Db 789 H-----KY--ILVSHSTNKEKILKO---YKITEEESKSLSCDPLD 824
Qy 799 YIF-----VYKGRNDAMDGYSV-----LYTRSAVL-----PEIIPELQTAQ 839
Db 825 LLENIGNNIPVMY---SMFDSLNNLSQLEMEIYEKEKVCYLKLNKNDIKNLEBAK 880
Qy 840 KVRGRDNTEIKTNTGPP-PLVERLEKKEVEGERT-----IIEVEEIT----- 884
Db 881 KV-----STVKTLSSSMQPLSLITQODKEFSANDTSHSTNLSNLSKLEFENILSLGRNK 936
Qy 884 -----EEVEKVRDEKVLFSKIFEFKELODEBNFLREKSEEMDVLDGLK 931
Db 937 NIYQELIGOKSSENFYEKELKDSDFEYNSEFNFYKSAADDINSLNDSKSK-----K 989
Qy 932 MEATEVEKLFGRALPIRKMAVAHCTFSPGDRIRFEFSSDDGIGRLGIRKIRNGTFL 991
Db 990 LE-EDINKL-----KTLQLSFDLYNKKYKLEERLEDKKRTGKYKMOIKKLT---LL 1038
Qy 992 KILPIQSADLRTGGRSSRP---LSAFRSGFSKGFIDIVLPKSKNELKELTAPLLKLV 1048
Db 1039 K-----EQLBSKLSLNNPKHVLQNSVFENK-----KKDAE----- 1071
Qy 1049 GVLACAFILVPSADAVALKTCACLKCCR---IELACIANPACAAVACLOTGNRPDE 1106
Db 1071 -----LAETENTLEMTKILKHYKGLVYNGESSPLKTLSEESIQTEDNVASL 1119
Qy 1107 TFCQIKCGDLFENSVDDEFNECAVSRKCPKRSKDLGEFAPDPVSVAOVFNISDRNGKW 1166
Db 1120 E-NFKVLSKLEGRKLDKNL---LEKKKLSLSSGLHLLA-EKEVKIKKNY----- 1167
Qy 1167 YITSGANFTFDFACQLEHFEFTEDDNKLVGNISWRIKTLDSGFTFSRAVOKEFOD----- 1222
Db 1167 ---TGNSSPS-----ENNTDVNNALE---SYK-KELPEGTDVAIVSESSDILEOS 1210
Qy 1222 -PNOGVLYNHNDNEYLHODDMYLLSKIKENKPEDIYFYVYRGNDK-WDYGGAAYVTR 1279
Db 1211 QPKRA-----STHVGAESNTITTSQNVDEVDVITVIFESSEBDYDGLQVY--TG 1262
Qy 1280 SSVLPNSIIPLEKRAKSGRDFSTFTRDNTGCEPALVERIEKTYEBSERTIIVEVEE 1339
Db 1263 EAVTPSVI-----DN-----ILSKIE---NEVEVLYLPLAG 1291
Qy 1340 IEEVEKEVEKVGRTMTLQRLAEGNEKODEBNFVRLSEKEMEFID 1389
Db 1292 VYRSIKOLE---NMVTFNVNVKDLNLSRPNREKFNKLESIDLIPYD 1338

```

```

RT heavy chain.*;
RL NUCLEIC ACIDS RES. 16:4737-4737(1988).
CC -I- FUNCTION: MUSCLE CONTRACTION.
CC -I- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -I- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -I- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-REPEAT REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -I- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LM)
CC AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE SPLIT FURTHER INTO
CC 2 GLOBULAR SUBSEGMENTS (S1) AND 1 ROD-SHAPED SUBSEGMENT (S2).
CC -I- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; L12104; G402372; -
DR EMBL; X07273; G49641; -
DR PIR; A28298; A28298.
DR HSP; P08799; LMND.
KW MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;
KW ATP-BINDING; METHYLATION; ALKYLATION; MULTIGENE FAMILY.
FT DOMAIN 1 838
FT DOMAIN 839 1934
FT DOMAIN 779 801
FT DOMAIN 839 1934
FT NP_BIND 177 184
FT DOMAIN 654 676
FT DOMAIN 756 770
FT MOD_RES 128 128
FT MOD_RES 694 694
FT MOD_RES 704 704
FT MOD_RES 966 966
FT CONFLICT 966 978
FT CONFLICT 986 986
FT CONFLICT 1008 1014
FT CONFLICT 1057 1057
FT CONFLICT 1060 1060
FT CONFLICT 1095 1095
FT CONFLICT 1217 1217
FT CONFLICT 1271 1271
FT CONFLICT 1327 1327
FT CONFLICT 1358 1358
FT CONFLICT 1504 1504
FT CONFLICT 1537 1537
FT CONFLICT 1556 1556
SQ SEQUENCE
1934 AA; 222928 MW; E586AA2F CRC32;

```

Query Match 188. Score 133.5; DB 1; Length 1934;
Best Local Similarity 17.0%; Pred. No. 6.3;
Matches 251; Conservative 211; Mismatches 525; Indels 493; Gaps 63;

```

Qy 188 NCGAVSRKKCYPRKSDVGEFVPDRN-----AVQNNMKMDESGKMYITSLGNTPT- 238
Db 390 NSADLKGKCHRVAVNGEYITKGNQVOVSYAALAKASYEKAFN--MMVTR-INAIL 446
Qy 238 -----FDAPDCQLEH---FHMENDKLVNLTWRITKTLDSGFTFSAVQ 277
Db 447 ETKQPROYFIVGLDIAGELIDFNSFEOLCINFTEKLOQFNNHVMVLEQEKYKKEGIE 506
Qy 278 -TFVO-----DPLPGALY-NHNDNEFLHYODM 303
Db 507 WTFIDFGMDLQACIDLKPKRIMSLIEECMPKATDMFKAKLDVNDLGNKSNFQKPR 566
Qy 304 YILSQINKEPDYITFYVYRGND---AMDGYGSGVITRSPLESTIPNLQKAA-KS 358

```



```

Db 567 NVKQKQAH-----FSLVHAGTVDYNIIGW-----LOKNKDPLENTVGLYOKSSKL 615
Qy 359 VGRDFNFTITDSCGPEPLVERLEKTAEGEKLIIKEAVEI--EEVEKEVEKVADE 416
Db 616 LSNIFANY-----AGADAP--VDKGRKAKKSSSF-----QTVSVLHRNKLKMTNLST-- 665
Qy 417 MTLFOLLEGEKLEQDEENFVELSKEKE-----LNELOEAT--EVEKLFGR 466
Db 665 -----HPHFRCIIPNETKSPGVMDNPLVMHOLRCNGVLEGIRICRG 707
Qy 467 LP-----IRKLMALAPHSNFLANHEIIKYVGSKLPGHKRFSGMEDYFGSI 515
Db 708 FPNRILGDFQRYRILNPAIPGOFIDSKGAEKLLSIDDHNOYKRGHTVFVKAG 767
Qy 516 VAKICSRIPRYFRKSPRICGGLDSRGLQLSHGKHLNLSPAHS--INONVPGNSGCK 572
Db 768 LGLLEHRDRERLRIITRI--QASRGLLSRMEFKKILERSDSLVIOWNI--RAFMGVK 824
Qy 573 -FP-----KDALMWEKMGQFAKTAIYATITLSVASKAADVALAKT 613
Db 825 NWPWKLYFKIKPLKSAETEKEKMATK--EDFGR-----VKDALEK 864
Qy 614 CTCLKEGRLELAKCISAPACANVACLOTNNRPDETCQIKCGDLFENSIV--VDE 668
Db 865 SEARKKELEEKVSLQE-----KNDLQVQAEQDNLADAEERDOLIKNKIQLEAKYKE 920
Qy 669 FNE-----CAVSRK--KCVPRKSDVGDVPVDPDSVLVOKFDMKDFSGKMFIT 713
Db 921 MTERLEDEEEMNAELTAKRKLEDCSELKRDIDLELTAKVEKDK----- 968
Qy 714 RGLNTPFAFCOLHEFTEENKLVGNLSWRIRTPDGGFFRSANOKVODPKYPIGLYN 773
Db 968 -----HATENK--VKNLTEEVAGLD-----ETIAKLTKEKALQEHQ 1003
Qy 774 HDNEVLLYODDY--ILSSKVE-----NSPEDIIFYVYKGRND--AMDGYGGSVLYTRS 823
Db 1004 QALDLOAEDEKVNLTLSKVKVLEBOVDLBESLQEKVRHDLERAKRKLEGLDKLQE 1063
Qy 824 AVLPEISIIPELOTAQKVG-----RDF-----NFIKIDNTCGPE-----PPLVERL 865
Db 1064 SIM-----DLENDKQODELKKKDFELNALNARIEDQALQSOLQKKLELQARIEL 1117
Qy 866 EKKVEGERT-----IKVEEIEEVEK----- 890
Db 1118 EEBL--EAERTARAKVEKRLSDLSRELETSERLEEGAGATSVQIEMNKKREAFPKMRD 1176
Qy 890 -----VADKEVTLFSKLFEGKELODEENFELSKREEMVLDGLKMEAT 935
Db 1177 LERATIQHEATAALAKRKHADSVALEGEOIDNLQRYKQLEKESEFKLEDD--VTS 1232
Qy 936 EYEKLFGRALPIRKLMVAVATHCETSPCHRIREFSSDDGIGRLGITRKINIFILKILP 995
Db 1233 NMEQIIRAKANLEKM-----CRT--LEDOMNEHRS-----RAEETORSVNDL----- 1273
Qy 996 PIGSADLRITGGSSRPL--SAFRSGSGKIPDIYVPLSKNELKELTAPLLKILGVLA 1052
Db 1273 TSORAKLOTEENGELSRQULDEKELISQULRG--KLIVYTOULDELRQOL----- 1319
Qy 1053 CAFLIWPSADAVDAKTCACILKGCRIELAKCIANPACANVACIOTCNNRDETE--CQ 1110
Db 1319 -----EEVAKAKNTLHAALQSAHRD-----CDLRLRQYEETEAKAE 1355
Qy 1111 IKC-----GDLEFNSVYDEFNECAVSRKCVPRKSDLGFFPAPDPVILVQN 1158
Db 1356 LQCVLSKANSEVAQWRTKJETDAIQRTTELEAKKKLAQLQDAEE----- 1402
Qy 1159 ISDFNGKWTITSGLPEDAFDCOLHEFTEGDNKLVGNISWRIKTLDGSGFFRSVOKF 1218
Db 1402 -----AVEAVNAKCSLEKTKHRLQNEIDLWVD-----VERSNAAA 1439
Qy 1219 VODPNQGVLYNHNDNEYLHYQDDMYIISKEIKPKPEDYITVYVYTRGRNDAMDYGGAIVIT 1278

```

```

Db 1440 AUDRKO-----ENFDKILAEWKQYEESSQSELESSQKE-----ARSLST 1478
Qy 1279 RSVLPNSITIPLE-----KAASIGRDFSTFIIRTDTCGPEDALYRIEKTVEEGRI 1332
Db 1479 ELFKLMAEESLEHLETFKRENNKNOLEISDLTEQLOSTKSHLEFKIKQLEAKME 1538
Qy 1333 IYKEVEIEEVEKVEKVEKVGRTMTLFO-----RLAEGNELKODENFVRELS----- 1382
Db 1539 LOSALEFNASJLHDEGNLNRQLEFNOIKAEIEKRLAEKODEMEOAKRNHLRVDSIQT 1598
Qy 1382 -----KEEMF--FLDEIKMAEVEKLEFKA 1405
Db 1599 SIDAETSRNEMALRYVKKMEGDLNEMEIOLSHANMAEA 1638

```

RESULT 14

```

DEF_BACAN
ID DEF_BACAN STANDARD: PRT: 809 AA.
AC P15917;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).
GN LEF.
OS BACILLUS ANTHRACIS.
OG PLASMID PXO1.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.
RX MEDLINE: 90034185.
RA BRAGG T.S., ROBERTSON D.L.;
RT "Nucleotide sequence and analysis of the lethal factor gene (lef)
RL from Bacillus anthracis.";
RN GENE 81:45-54(1989).
RN [2]
RP SEQUENCE FROM N.A.
RA LOWE J.;
RL SUBMITTED (APR-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP ZINC-BINDING.
RX MEDLINE: 95154669.
RA KOCHI S.K., SCHIAVO G., MOCK M., MONTECUCCO C.;
RT "Zinc content of the Bacillus anthracis lethal factor.";
RL FEMS MICROBIOL. LETT. 124:343-348(1994).
CC -1- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. LF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED
CC WITH PA, CAUSES DEATH. LF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO
CC BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY
CC FACILITATING THE INTERNALIZATION OF LF OR EF.
CC -1- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT
CC PROTEINS, A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN
CC EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B. ANTHRACIS EF
CC AND LF.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC
CC METALLOPROTEASES).

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M29081; G143144; -
 EMBL: M30210; G143142; -
 PIR: J00032; J00032.
 PROSITE: PS00142; ZINC_PROTEASE_1.
 HYDROLASE; METALLOPROTEASE; ZINC; TOXIN; SIGNAL; REPEAT; PLASMID.

FT SIGNAL 1 33
 FT CHAIN 34 809
 FT DOMAIN 34 293
 FT DOMAIN 300 420
 FT METAL 719
 FT ACT SITE 720 720
 FT METAL 723 723
 SEQUENCE 809 AA, 93786 MW, DB1B6EBB CRC32;
 LETHAL FACTOR.
 PA-BINDING REGION (POTENTIAL).
 REPEATS.
 ZINC (CATALYTIC) (POTENTIAL).
 POTENTIAL.
 ZINC (CATALYTIC) (POTENTIAL).

Query Match 1.8%; Score 132; DB 1; Length 809;
 Best Local Similarity 17.2%; Pred. No. 2.4;
 Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 EENKLVGNSMRTIPDGGFFTRSAVOKFVODPKYPGILYVND--NEYLLYODDWY--I 787
 DB 105 EYKRAIG--KIYVDGDTIKHLSLEALSEDKKIKINDYKDALLHHVYAAKEGYEPV 161
 QY 788 LSSKVENSPEDYIFVYVYGRNDAMWGCGSVLYTRSAVLPESITPELQTAQKGRDFNT 847
 DB 162 L---VIGSEEDYVENTEKALN-----VYIEIGKILSDILSKINOPYOKELDVLT 209
 QY 848 FIKTNTG-----PEPPLVERLEKVEEGERTIK----- 879
 DB 210 IKNASDSQODLFTNQLKEHTDPSVFLQNSNEVQVAFKAFATYIEPQHRDVLYQLY 269
 QY 879 -----EVEIEEVEKVRKDEYTLFSKLEPFGEKE----- 908
 DB 270 APEAFNWDKNEOEINISLELKDORMLSRVFEKWEKIKQHYQWSDSLSEGRGLLKL 329
 QY 908 ---LQREBNFLRELSKEMDVLDGLKWEATEV---EKLFGRALP-----RKL 950
 DB 330 QIPIPKKDDIITHSQEKEKELKRIQDSDPLSTEKEFLKQLQDIRDSSEKEKL 389
 QY 951 MAVATICTSPCHDRIRFESSDDGIGRLGTRKRINGTFLKILPQISAD---LRTTG 1006
 DB 390 L-----NRIQVSSNP-----LSEK--EKEFLKLLKDIDQPIQDRIQRLDITG 429
 QY 1007 GRSSNPL-----SAFRSGFSKGIPIVLPFSKNEKELTAPLLKLKLVG 1049
 DB 430 GLIDSPSINDLVKQYKRDIONIDALLHOSIGSTLYNKRIYLYENMMINMLTATL----- 484
 QY 1050 VLAQAFILVPSADAVDALKTCACLLKGCRIGELAKCIANPACANVACLOTCNNRPDETC 1109
 DB 484 -----GADLVST----- 492
 QY 1110 QIKGDLFENSVD--FNECAVSRKRCVPRK---SDLGEFPAPDPSVLVONENISDENG 1164
 DB 492 -----DNTRKINGLIFNEFKKFKYSSNYMIVDINERPALDNERLKWRIQSPDTR 543
 QY 1165 KWTYTGILNPTFADFQDLHFHTEGDNKLY--GNISMRKILDSGFFTRSAVQKVDPP 1222
 DB 544 AGYLENG-----KILQIRNIGLEIKDVO----- 567
 QY 1223 NQPGVLYNHDXLHYODDWYILSKLENKPEYIFVYVYGRNDAMWGYGAVVYTR--S 1280
 DB 567 -----LIKQSEKEYIRI-DAKVVPKSKIDKIQE---AOLINQENKRLGLPKTKLIT 617
 QY 1281 SVLPNSITPLEKAASIGDESTFRTDNTGCPALVERIEKTEVEGE-RIIVKEV-- 1338
 DB 618 FVNHNRVYASVIVLILNEMNNIOSD-----LIKVTNYLVQNGRFVFTDITL 669
 QY 1338 -----EELIEEVEKVEVYGRTEMTLFORLAEGNELKQDENNVRE----- 1380
 DB 670 PNIAEQYTHQDELHYOVHSGLYVPESRSLILHGPBGV--ELRNDSEGHIEFGHAYVDY 728
 QY 1380 -----LSKEEME-----FLDETKEMASVY 1398
 DB 729 AGYLLDKNOSDLVYNSKRFIDIFKEGNSUL 758

RESULT 15
 Y109_YEAST

ID Y109 YEAST STANDARD; PRT; 1679 AA.
 AC P40457;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
 DE HYPOTHEICAL 195.1 KD PROTEIN IN DNA43-UB11 INTERGENIC REGION.
 GN Y1149C.
 OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA, FUNGI, ASCOMYCOTA, HEMIASCOMYCETES, SACHAROMYCETALES;
 OC SACHAROMYCETACEAE, SACHAROMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
 RA CHURCHER C.M., CONNOR R., CORSEY T., DEAR S., DEVLIN K., FRASER A.,
 RA GENTLES S., HAMLIN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
 RA LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,
 RA RAJANDREAN M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
 RA WALSH S.V., WHITEHEAD S.;
 RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; 247047; G763197;
 DR EMBL; 238059; G557774;
 DR PIR; S48385; S48385.
 KW HYPOTHEICAL PROTEIN.
 SQ SEQUENCE 1679 AA, 195141 MW, 5897CD94 CRC32;

Query Match 1.7%; Score 130; DB 1; Length 1679;
 Best Local Similarity 17.9%; Pred. No. 8.3;
 Matches 210; Conservative 188; Mismatches 396; Indels 376; Gaps 59;

QY 379 LVERLE-----KTAESEKILLKEAVEIEEVEKVEKVDTEMLTFLRGLGFEQLQ 431
 DB 536 LADLKEYGKODKTLQKVENQTIKEAKDAIIELE-NINAKMETRINILLEROSYKLA 594
 QY 432 QDEENFVR-----ELSEKEKILLNLOMATEVEVKLFGALPIRKR----- 474
 DB 595 STEENKANNINSVTSMFAANEKIRLELELSSTVE---NSAIIQNLKELLIIYKRSOC 650
 QY 474 ---MALAPHSN--LAN-----HETIRYVGSKLPGHRSFGMEDYFGSIYVAKICSS 522
 DB 651 KKKTLLEDFENFKGLAKEKRMLEAIDHL--KALELKQSW----- 691
 QY 523 RRIPIRFR-KSPRIGCGDLSRGLOFSGKHNLSPASHINON-VPKNS-----GCKF 573
 DB 691 ---VPSTIHEKRPASTELSQSRIKISL-EYEISKLKTEKTSFIPRESLTPDEQCKE 747
 QY 574 PKDVALMVE-----KMGOFATATVAI-----FILVASKADAVALKT 613
 DB 748 KKLQRLKLEKESLISHNENKMPSSKEGYO-KAKIKELNNLERLSQLOSKOIEIESIRS 806
 QY 614 CTCLKECFLLELAKTISNPACANVACLOTCNNRPDETECOIKGDLFENSVDVEFECA 673
 DB 807 C---KDSQLKWAQ-----NTIDTEKMK-----SLTLELSNKE 837
 QY 674 VSRKKCVPRKSPVGFPPDPSPVLVQKFDKMDGSGKFTIRG---LNP-----FDAED 724
 DB 838 TTIIEKI---SSIEINL---DKELRKTFRQYK-----FLQNSDPASTLIEFTLKRKELEQIO 885
 QY 725 COLHEFHTENKLVGNLSMRIRTPDGGFFTRSAVOKFVODPKYPGILYVNDNRYLYLQXD 784
 DB 886 VOLKDANS-----QIQAIEEIISSNENALI----- 911
 QY 785 WYILSSKVENSPEDYIFVYVYGRNDAM-----DGYGGSVLYTRSAVLPESITPELQ 835

```

Db 911 --ELKNEELAKTRENDAKIELEKKEMAREEDLSRLGELGEI-----RALOPKX 959
QY 836 TAAQVGRDENTFTDNTCGPEPLVERLEKKEVEGER--TII-----KEVEEIEEVE 888
Db 960 EGALH-----FYQOSEKLRE--VERIQMIKIEKMSIYOLCKKEMSQYOSTMK 1009
QY 889 KVRD-----KEVT-LFSKLFEGFELQDEENFLRELSEEMVDLQK 931
Db 1010 ENKDLSELVIRLEKDAOCQELITKSSLYSAODLLDKHERKMEKADYERELISNIE 1069
QY 932 M-EATEVEKLGRAPIRKIMAVATH-----CFTSPCHDIRIFFSSDDIGRLG 979
Db 1070 QTESLRVE---NSVLIKVDOTLANNQDKHLKVSLSFNLHER-----1112
QY 980 ITRKRINGTFLIKIIPLOSADLR--TTGGRSSRPLSAFRSGFSKIFDVIPLPSKNEKE 1038
Db 1112 -----NSLETKLITCKRELAFVKQKNDSEKTIINDLORTOTLSE-KE 1152
QY 1039 LTAFLKLKVLGACAFILVPSADAVDALKTGACLLKGRLELAKCIANPACAAVACLQ 1098
Db 1153 -----YQCSAVIIDEFKDITKEVTOVNLKENNALLOKSLKN-VTEKNREIYK 1199
QY 1099 TCNNRPDETECOIKCGDLFENSVDENECAY----SRKCCVPRKSDLG--EFPAPDP 1151
Db 1200 QLNDRQOEISRLQR--DLIQTKQVINSNKLIVYESEMECKORYODLSQOQKDAQKD 1257
QY 1152 VLVQNFENISDFNGKWTITSGLNPTDAPDCQJHEFTEBGNKLVGNISMRIKITLDSGFT 1211
Db 1258 IEKLTNEISDLKG--LSSAENANADL-----ENKFN-----RLK-----1291
QY 1212 RSAVOKFVODPNQPGVLYNHDNEY-----LHYODWYI--LSKI--ENKPEDYI 1257
Db 1291 KOAHEKLDASKKOQALTELNELKAIKDKLEODLHFENAKYIDLDTIKAKAHELOSEDV- 1350
QY 1258 FVYYRGNDAMDGYGAVVYTRSSVLPNSIIPLEKAAKSIGRDFSTFIIRDNTGCPPEPA 1317
Db 1350 -----SRDHEKDTY-----FTLMEIE--SLKKELOIF---KTANSSSD 1383
QY 1318 LVERIKETVE-EGERTIVEVEEIEVEVEKEVEKVGRTENTLFORLAEGFNE-LKODEEN 1375
Db 1384 APEKLVNMEKEKDRILIDERTKEFEKLIQETLTKSTSSAAYSKDIEITLKKEMLKEYEDE 1443
QY 1376 FVRELSEEMEFLDEIKMEASE--VEKLFQK 1404
Db 1444 TLRRIKEAENLKKRIRLPSERIOKITISK 1473

```

Search completed: October 14, 1999, 03:58:19
Job time: 2412 sec

1	2557	34.1	478	10	040593	040593 nicotiana t
2	2508	33.5	473	10	040251	040251 lactuca sat
3	2430	32.4	462	10	039249	039249 arabidopsi
4	194.5	2.6	2269	5	026223	026223 plasmodium
5	167.5	2.2	2401	5	026216	026216 plasmodium
6	156	2.1	1365	2	046525	046525 mycoplasma
7	156	2.1	1933	13	090337	090337 cyrinus c
8	155.5	2.1	886	1	028230	028230 archaeeoglob
9	151	2.0	1939	5	025662	025662 plasmodium
10	147	2.0	3113	4	013246	013246 homo sapien
11	146	1.9	2748	3	003767	003767 saccharomyc
12	144.5	1.9	839	5	026024	026024 plasmodium
13	144	1.9	2166	2	051465	051465 borrelia bu
14	143.5	1.9	1558	5	096275	096275 plasmodium
15	142.5	1.9	1819	2	092LVO	092LVO helicobacte
16	142.5	1.9	1139	5	P91349	P91349 caenorhabdi
17	142.5	1.9	5105	5	061201	061201 caenorhabdi
18	141	1.9	1002	2	066583	066583 aquifex aeo
19	140.5	1.9	1302	2	049547	049547 mycoplasma
20	140	1.9	800	1	059066	059066 methanococc
21	139.5	1.9	1025	3	012176	012176 saccharomyc
22	137	1.8	1109	6	000756	000756 cryotolajus
23	137	1.8	1676	10	023332	023332 arabidopsi
24	136.5	1.8	3724	5	077320	077320 plasmodium
25	136	1.8	1137	12	039YV6	039YV6 melanoaplus
26	135.5	1.8	1312	4	092878	092878 homo sapien
27	134.5	1.8	946	2	P70888	P70888 bacteroides
28	134	1.8	978	2	067124	067124 aquifex aeo
29	134	1.8	620	4	043663	043663 homo sapien

30	261175	0621175	caenorhabditis
31	134	090339	cypripus ca
32	133.5	043254	homo sapien
33	133	P91121	caenorhabdit
34	133	0961133	plasmodium
35	132.5	066878	aqifex aeo
36	131	028714	archaeoglob
37	130.5	049546	mycoplasma
38	130	028641	oryctolagus
39	130	042352	cypripus ca
40	129.5	048673	lactococcus
41	129.5	1992 13	xenopus lae
42	129.5	042263	xenopus lae
43	129	059462	pyrococcus
44	129	018392	drosophila
45	129	2919 12	rice stripe

RESULT	1			
040593				
ID	Q40593	PRELIMINARY;	PRT;	478 AA.
AC	Q40593;			
DT	01-NOV-1996 (TREMBLrel. 01.	Created)		
DT	01-NOV-1996 (TREMBLrel. 01.	Last sequence update)		
DT	01-JAN-1999 (TREMBLrel. 09.	Last annotation update)		
DE	VIOXANTHIN DE-EPOXIDASE	PRECURSOR.		
GN	TYDEL1.			
OS	Nicotiana tabacum (Common tobacco).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta			
OC	euclaryophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;			
OC	Nicotiana.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=XANTHI; TISSUE=LEAF;			
RA	BUGOS R.C., YAMAOTO H.Y.;			
RL	Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; U34817; AAC50031.1; -			
DR	MENDEL; 9222; NTCa.Vdel.1.			
KW	Transit peptide.			
FT	TRANSIT	1	134	POTENTIAL.
FT	CHAIN	135	478	POTENTIAL.
SQ	SEQUENCE	478 AA.,	54561 MW,	9582AD42 CRC32;

Query Match	34.1%;	Score 2557;	DB 10;	Length 478;
Best Local Similarity	100.0%;	Pred. No. 2.5e-139;		
Matches 477;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	474	MALAPSHNFTAAHETIKKYVSGKLPBGHRRPSWGMEDYFGSIVYAKICSSRRIPRYFRKSP	53
Db	1	MALAPSHNFTLANHETIKKYVSGKLPBGHRRPSWGMEDYFGSIVYAKICSSRRIPRYFRKSP	60
Qy	534	RICCGDLSRGLCTFSHGKHNSPAHSINQNPKNSSCKEPRKDVALMWEKKGOFAKTAI	59
Db	61	RICCGDLSRGLQTFSHGKHNSPAHSINQNPKNSSCKEPRKDVALMWEKKGOFAKTAI	120
Qy	594	VAFIITSVASKDADVALKTCCTCLLKEBRLFLAKCISNPACANAVACLOTQNNRDETEC	65
Db	121	VAFIITSVASKDADVALKTCCTCLLKEBRLFLAKCISNPACANAVACLOTQNNRDETEC	180
Qy	654	QIKRGDLEFNZYVDFNEECVARRKCYPRKSDVDPEVPSPSYLVQKFKMDFSKMFIT	71
Db	181	QIKRGDLEFNSVYDFNEECVARRKCYPRKSDVDPEVPSPSYLVQKFKMDFSKMFIT	240
Qy	714	RGLNPTFDADFQOLHEFHTEENKLVGLNSWRIRTPDGGFFTRSAVQKFVQDEKRYGILYN	77
Db	241	RGLNPTFDADFQOLHEFHTEENKLVGLNSWRIRTPDGGFFTRSAVQKFVQDEKRYGILYN	300
Qy	774	HDNEYELLYQDDWYILSSKVENSPEDYITVVIYKGNDAAMDYGGSVLYITSANVLPSIILPE	83

```

Db 301 HDNEYILLYQDDWYIILSSKVENSPEDYIFVYIKRNDAMOGYGSVYITTSALVPESTIPE 360
QY 834 LQTAAGKVGADNFNTIKTNTGCGPEPLVERLEKKEVEEERTIIKEVEEIEEVEKVRDK 893
Db 361 LQTAAGKVGADNFNTIKTNTGCGPEPLVERLEKKEVEEERTIIKEVEEIEEVEKVRDK 420
QY 894 EVTLFSKLFEGFELQDEENFLRELISKEEMDVLDGLKMEATEVEKLFGRALPIRKL 950
Db 421 EVTLFSKLFEGFELQDEENFLRELISKEEMDVLDGLKMEATEVEKLFGRALPIRKL 477

RESULT 2
ID 040251 PRELIMINARY: PRT: 473 AA.
AC 040251:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JAN-1999 (TREMblrel. 09, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN VDEL.
OS Lactuca sativa (Garden lettuce).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Asterales; eustersids II; Asteraceae;
OC Lactuca.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ROMAINE.
RX MEDLINE: 96270536.
RA BUGOS R.C., YAMAMOTO H.Y.;
RT "Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
RT and expression in Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 93:6320-6325(1996).
DR EMBL: U31462; AAC49373.1;
DR MENDEL: 8691; LACsa.vdel.1.
KW TRANSIT peptide.
FT CHAIN 126 473 POTENTIAL.
FT SIGNAL 125 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 473 AA; 54447 MW; 102E7001 CRC32;

```

Query Match 33.5%; Score 2508; DB 10; Length 473;
 Best Local Similarity 100.0%; Pred. No. 1.6e-136;
 Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MALSLHTVFLCKEALNLYARSPCNERFHSQOPPTNIIMKIRSNNGFNSFLFTSYK 60
Db 1 MALSLHTVFLCKEALNLYARSPCNERFHSQOPPTNIIMKIRSNNGFNSFLFTSYK 60
QY 61 TSSFSDDSHCKDKSQISIDTSFEETIQRDLKRGMTLLLEKQWFOIATVLCVETVY 120
Db 61 TSSFSDDSHCKDKSQISIDTSFEETIQRDLKRGMTLLLEKQWFOIATVLCVETVY 120
QY 121 PRVDAVDAKTAACILKERIRLACIANPSCAANVACLOTNNRPDETQIKCGDLFE 180
Db 121 PRVDAVDAKTAACILKERIRLACIANPSCAANVACLOTNNRPDETQIKCGDLFE 180
QY 121 PRVDAVDAKTAACILKERIRLACIANPSCAANVACLOTNNRPDETQIKCGDLFE 180
Db 121 PRVDAVDAKTAACILKERIRLACIANPSCAANVACLOTNNRPDETQIKCGDLFE 180
QY 181 NSVVOFNECAVSRKCKVPRKSDVGEFPYDPDNNAVONFNMKDFGKMYITSGLNPTFDA 240
Db 181 NSVVOFNECAVSRKCKVPRKSDVGEFPYDPDNNAVONFNMKDFGKMYITSGLNPTFDA 240
QY 241 FCCQJLHEFHENDKLVNTWIRIKTLDDGFTFSRVAQTFVDDPLGALYNHDEFLHYQ 300
Db 241 FCCQJLHEFHENDKLVNTWIRIKTLDDGFTFSRVAQTFVDDPLGALYNHDEFLHYQ 300
QY 301 DDMYLLSSQIEKKPDYIFVYIRGNDAWDGSGSVITRBSPTLPESTIIPMLQAAKSVG 360
Db 301 DDMYLLSSQIEKKPDYIFVYIRGNDAWDGSGSVITRBSPTLPESTIIPMLQAAKSVG 360
QY 361 RDNFNFITDNSCGPEPLVERLEKTAEEGKLLKEAVEIEEVEKVEKVRTEMTLF 420
Db 361 RDNFNFITDNSCGPEPLVERLEKTAEEGKLLKEAVEIEEVEKVEKVRTEMTLF 420

```

```

QY 421 ORLIEGKEQLQDEENFRELISKEEKLINLOMEATEVEKLFGRALPIRKL 473
Db 421 ORLIEGKEQLQDEENFRELISKEEKLINLOMEATEVEKLFGRALPIRKL 473

RESULT 3
ID 039249 PRELIMINARY: PRT: 462 AA.
AC 039249:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-JAN-1999 (TREMblrel. 09, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN AVDEL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA.
RX BUGOS R.C., YAMAMOTO H.Y.;
RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA SHINN P., BUEHLER E., DEMAR K., FENG J., KIM C., LI Y., SUN H.,
RA CONWAY A., CONWAY A., KURTZ D., OUI O., SHEN Y.K., TORIUMI M.,
RA VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A., THROLOGIS A.,
RA ECKER J.R.;
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: U4133; AAC5032.1;
DR EMBL: AC003981; AAC14029.1;
DR MENDEL: 6341; ARATH.vdel.1.
KW SIGNAL.
FT CHAIN 114 462 POTENTIAL.
FT SIGNAL 113 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 462 AA; 52017 MW; 37FB33BA CRC32;

```

Query Match 32.4%; Score 2430; DB 10; Length 462;
 Best Local Similarity 100.0%; Pred. No. 4.5e-132;
 Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 951 MAYATHCFTSPCHDRTRFSSDDGIGRLGTRKRNGTFLKILPPIOGADLRITGGSS 1010
Db 1 MAYATHCFTSPCHDRTRFSSDDGIGRLGTRKRNGTFLKILPPIOGADLRITGGSS 1010
QY 1011 RPLSARFSGSKGIFVPLPSKNEKELTAPLLKIVGLACAFIYPSADVALKTC 1070
Db 1011 RPLSARFSGSKGIFVPLPSKNEKELTAPLLKIVGLACAFIYPSADVALKTC 1070
QY 61 RPLSARFSGSKGIFVPLPSKNEKELTAPLLKIVGLACAFIYPSADVALKTC 120
Db 61 RPLSARFSGSKGIFVPLPSKNEKELTAPLLKIVGLACAFIYPSADVALKTC 120
QY 1071 ACLKSCRIELAKCIANPACAAVACLOTNNRPDETQIKCGDLFENSVDNECAV 1130
Db 1071 ACLKSCRIELAKCIANPACAAVACLOTNNRPDETQIKCGDLFENSVDNECAV 1130
QY 121 ACLKSCRIELAKCIANPACAAVACLOTNNRPDETQIKCGDLFENSVDNECAV 180
Db 121 ACLKSCRIELAKCIANPACAAVACLOTNNRPDETQIKCGDLFENSVDNECAV 180
QY 1131 SRKCYPRKSDGGEFPAPDPVLYONFNISDPNGKMYITSGLNPTFDAFCQJLHEFHREG 1190
Db 1131 SRKCYPRKSDGGEFPAPDPVLYONFNISDPNGKMYITSGLNPTFDAFCQJLHEFHREG 1190
QY 181 SRKCYPRKSDGGEFPAPDPVLYONFNISDPNGKMYITSGLNPTFDAFCQJLHEFHREG 240
Db 181 SRKCYPRKSDGGEFPAPDPVLYONFNISDPNGKMYITSGLNPTFDAFCQJLHEFHREG 240
QY 1191 DNKLGNISMRITKTDGFTFSRVAQTFVDDPLGALYNHDEFLHYQ 1250
Db 241 DNKLGNISMRITKTDGFTFSRVAQTFVDDPLGALYNHDEFLHYQ 1250
QY 1251 NKPEDIIFVYIRGNDAWDGSGSVITRBSPTLPESTIIPMLQAAKSVG 1310
Db 301 NKPEDIIFVYIRGNDAWDGSGSVITRBSPTLPESTIIPMLQAAKSVG 1310
QY 1311 TCGPEPALVERLEKTAEEGKLLKEAVEIEEVEKVEKVRTEMTLF 1370
Db 361 TCGPEPALVERLEKTAEEGKLLKEAVEIEEVEKVEKVRTEMTLF 1370

```

QY 1371 ODENFVRELSEKMEFIDEIKMEASEVEKLEFGALPIRKVR 1412
 DB 421 ODENFVRELSEKMEFIDEIKMEASEVEKLEFGALPIRKVR 462

RESULT 4
 Q26223 PRELIMINARY: PRT: 2269 AA.
 AC Q26223:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, last annotation update)
 DE RHOPTRY PROTEIN
 OS Plasmodium berghei yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YM:
 RX MEDLINE: 95021522.
 RA KEEN J., SINHA K., BROWN K., HOLDER A.:
 RT "A gene coding for a high-molecular mass rhoptry protein of
 RT Plasmodium yoelii."
 RL Mol. Biochem. Parasitol. 65:171-177(1994).
 DR EMBL: 127838; AAA21304.1;
 SQ SEQUENCE 2269 AA; 265158 MW; E0A79FA6 CRC32;

Query Match 2.6%; Score 194.5; DB 5; Length 2269;
 Best Local Similarity 18.4%; Pred. No. 0.0069;
 Matches 255; Conservative 213; Mismatches 436; Indels 483; Gaps 65;

QY 304 YIISSQIENKPPDIYFVYRGNDAMDGYGGSVITRS-----PLIPESIIPNIO--KA 355
 DB 34 YIISNQIKKLN--VSTYEGR---EGFTSLELAKSEKTKLETITELTSSNEETVRL 87
 QY 356 AASVGDFFNNFTITDSCGPEPPLVERLEKTAEEGKL-----LIKEAVEI 401
 DB 88 EKIRIRLFKKY-----LDEAEARKYLEGLKLELNKKIKIDIAKIEYKATVVL 135
 QY 402 EEEVEE-----EYKAVDTMTLFLRLEEFKELOO--DEENFRELSEK 446
 DB 136 KKEIEKNNAVYDELANQSPKYVGIENKNTIYNTIKSYEDQIYEGDIDTFYNELSIVK 195
 QY 447 E-IILNLOEATVEKLEFGALPI--RKLMAIAPHSNLANET-----IKYV 493
 DB 196 EDPIDIE--DKTKLEMLRSKIDNVYDKIQMEIETVYKSHNINETNNKIPNILEIKYI 254
 QY 494 GSKLPGRKRSWGMEDYFGSIVAKICSSRRIPRYFRKSPRICCGDLSRGLQLEFHH- 550
 DB 255 YDEI--SKELNKLDEPFKNK---EKELSNKISDYDKKREQL--SEYKSKMLEIRHYNQ 307
 QY 550 -----GKNLSPAHINONV-----KGNSGCRPKOV 577
 DB 308 TYNVNTKEEBAKQNYDKSNHEMTIIPNEDEISKIIEVATMKDELISKVNTIDFNKKY 367
 QY 578 ALAWWEKGOFAKTAIYAIPIILSVASKADAVDAKTKTCLLKECRLELAKCISNPACAA 637
 DB 368 KEYVNSHSHOFT-----LTDKIKAEVSDK-----ELKK----- 397
 QY 638 VACLQTCNNRP--DETEQIKCGDLEFN-----SVDEFNECAVSRKKCVPRKSD----- 686
 DB 397 --CQSFNDKNSLINTKNSIE--KEYQNTINTLKVDEYIKVCKSTKESITKFSQOTII 452
 QY 686 -----VGDPVPDPSVIVOKFDMKDFSGKWFITRRLNTPFAFDQQLHEFHHE- 735
 DB 453 KDMLNQIKTYKETSNDKS--YIEKEF--QILITGR--QTLENKFTIEFSLNHEANNEL 507
 QY 735 ---NKLVNLSWRIRTPDGGFFTRSAVOKFVODPKYFGILYNNHNDYLLYODWYIIS 790
 DB 508 IKYISDLKANGINEMMLYNQIFEK--EKTFNIDKEKNI--HINEI-----S 552
 QY 791 KVENSPEDYIF-----VYKGRNDAMD-----GY 814

DB 553 KIEIKHASIVNISEETEREIGINIESLNTVKEVKNVNLNKKIKELKHYSDFGK 612
 QY 815 GGSVLYTRS-----AV-----LPSIIPETQTAOKRGDFNTFI 849
 DB 613 EGNIKYTDKIKKINDDIKAVSOQIDOHINGLDDIDOKSSSESYSEKKEQINKLEKYSNEI 672
 QY 850 KIDNTCG---PEPPLVERLEK--VEEGERTIIEVEEIEE--VEKVRDEVTLEFSLK 901
 DB 673 SNQNVGIRKKQOIIYTKIDKKNIYEEINKLLEISIKTEKDNSTLEKVDINLSYGQNL 732
 QY 902 FEEFKELQDNEFNFLPELSKEH-----DYDGLKMEATVEKLEFGALPIRKLMV--AT 955
 DB 733 GNLFLE-QIDEEKKKAENNTIKSMEAYIDDLIDNKKRSQEIETMDIKMDINKEMALKIS 791
 QY 956 HCTSPCHDIRFESSDDGIGRLGTRKRLNGFFLLKILPIQSADLFTGGRSSRPLSA 1015
 DB 792 HDDDKCHDKSK--NHKENISDIYDKSSKIQDF-----SRESINDIKKILKNVSE 842
 QY 1016 FRSGFS-----KGIDVPLPSKNELKEL-----TAP 1042
 DB 843 SQNHNSINDQCIIEVANITYNIIKL--NKIKITIDRYKEVTESEKKNKINDENLNSSEK 899
 QY 1043 LILKIVGLACAPLIYPSADAVALATCACLKGCRIELAKCIAN--PACANVACIOT 1099
 DB 900 VIKKIEGDL-----LKECRSKINSTLDDKDIDECIKINIVLKK 938
 QY 1100 CNRPDETEC--QIKCGDLFENSVDENEC----- 1129
 DB 939 -NINNETNITNHFKNAEIKVLSNFNNIEMADKSOYILEKKNNGTNDHDYNIKEL 997
 QY 1129 -----AVSRKCVPRKSDIGFPPADPSVLYONENISDPNGKWYITTSGLNFT 1175
 DB 998 KSHDKNGKYTEADQKKAIOKKLEFEOYKEEVVILLKRYAVELAKN- 1048
 QY 1176 FDADQCQLEHRTGDKKLVGNISWRK-----TIDSGFFTRSAVOKFQDNPQVLY 1229
 DB 1048 --FD-----KTKNDSK--QIITKIDAHNYCTLESQ-----KSKKKNNE- 1083
 QY 1230 NHDNEYLHYODWYILSSKIENKPEDIYFYVYRGNDAMDYGAVYTRSSVLP----- 1285
 DB 1083 -IKNEKIHIDE--VANNDKSNK-----AITSIKVSEPEPKTI 1118
 QY 1285 --NSI-----IPELEKAASIGRDSFTIRIDNTCGEPALVERIEKTEVEGER 1331
 DB 1119 IKINEIRFSDCLKETNDEKQISNLSIDQETKLTEN--GQOLKTELELSKKOK- 1176
 QY 1332 IIVKEVEIEEFVAKVEKVAQRTMTLFOR-----LAEGFNELKODEENFV--RELS 1381
 DB 1176 --KNIEDQKEDEVNSKININENTVNOHKKNYEIOIVEKINEIAKTNNQIESTRELI 1232
 QY 1382 KEEMFEL 1389
 DB 1233 KPTIOHI 1239

RESULT 5
 Q26216 PRELIMINARY: PRT: 2401 AA.
 AC Q26216:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, last annotation update)
 DE RHOPTRY PROTEIN (FRAGMENT)
 OS Plasmodium berghei yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YM:
 RX MEDLINE: 9707745.
 RA SINHA K.A., KEEN J.K., OGUN S.A., HOLDER A.A.:
 RT "Comparison of two members of a multigene family coding for
 RT high-molecular mass rhoptry proteins of Plasmodium yoelii."
 RL Mol. Biochem. Parasitol. 76:329-332(1996).

DR EMBL: U36927; AAB41263.1; -
 FT NON-TER 1
 SQ SEQUENCE 2401 AA; 281980 MW; 39A40760 CRC32;

Query Match 2.2%; Score 167.5; DB 5; Length 2401;
 Best Local Similarity 15.7%; Pred. No. 0.26; Indels 575; Gaps 71;
 Matches 260; Conservative 274; Mismatches 498;

```

QY 23 PCNERFHSGQPTNTIMMK--IRSNNGYFNSFRFTSYKTSFSDSSCHKDKSQICSID 80
DB 453 PENMEKYQ--KPSIEIKIKKDEFELSKVKNKYNDFD--KYKKEVSEHNKFTLETAKITE 508
QY 81 TSEFEORFEDLRGMWMLLEKOWROFIQALVCTEYVPRVDAVDALKTACILKECR 140
DB 509 VDEDELEKTEKFNDSKSLINETKKSIEEYONINT--LKKVD--DYIKVC----- 556
QY 141 IELACIANPSCAANVACLOTGNNR--PDETECOIKC-----GDIENSVD 185
DB 556 LMTNELITN-----CHNKOTILKDLNINIKITKETSIDKTYDKENITLD 603
QY 186 Q-----FNECAVSRKCV--PRKSDVGEFPVDRNAVONNMKDFSGKMY 229
DB 604 KTELEIKFTGLSLNNHSSNNKELLTYFYDLKANLG--KKNMLYKQFNEKEKAVE-- 659
QY 230 ITSGINFTFADPCOLHEFHMEHNDKLVGNLWRIKTLDSGFTFRVAVOTFVODDLRAL 289
DB 659 -----DIKKKNDINKIVSNIEITYT-----SI 682
QY 290 YNHDN-----EFLHYODWYILS-----SOIENKBDYLFYVYGRNDAMD 330
DB 683 YNINDTENEIGKSIELINTK--VLEKRVANVTNLEIKELKDXDFOF----- 731
QY 331 GYGGSVITRSPILPESILPNLOKAKSVGRDNFNFTTDSGCEPPEYERLEKTBEG 390
DB 731 GKEKNIKIPDENKIKNDIDTLNOKDKSIEITLTKKSNHMDIEIKQIDKLKV--PN 788
QY 391 EKLKKEAVEIEEVEKEVEKVRDTEMLFO--RLLEGFEKLODEEN----- 437
DB 789 KTFMNEDEKREIEKLENIIVEKI--DKKNITYKEIDKLEINIEKINDTSLKLNINLSY 847
QY 437 -----FYRELSKEEKEI-----LNELOMEATEVEKFLGRALPI-----RK 471
DB 848 GKSIGNLELOQIDEKKAHEHTIKAMEAYIDDLNKKSOIEKKEKNINIMDKIMDKHE 907
QY 472 LRNALAPHSNFLANHETIKYVGSKLPGHKRFS-----MGWEDYFGSI-----YAK 518
DB 908 MKALINISHDYKIKYHTSKNH--EKKISDIRKNSIKIIOFSEESIINDIKKELEKNVLES 966
QY 519 ICSSRRIPRYFRKSPRICCGGLDSRGLOLFSGHNLSPAHISINOVKGNSGCKFPKDAVA 578
DB 967 QNNNTDINQYLSKIENTI-----YNIILKLNKIKIIDK----- 999
QY 579 LMWEEKMGOFKTAIVAFILTSVASKADAVDALKTCTCL-----LKRCLLELAKCISN-- 632
DB 999 --VKEYTDEIEKN--NKKINMELNSKEIITOLKENSLSKCCOSKIKSTIDNNT 1048
QY 632 -PACANVACLOTG--NNRPDETCQIKCGDLFENSVDENECAVSRKCVPRKSD 685
DB 1049 VSECIKNTINLKIYIVNEKN--INTYFKNA--EYINQ----- 1083
QY 686 VGDFPVDPVSLVQKPMKDFSGKMFITRGLNPFPAFCOLHEFHTEENKVLGNISWRI 745
DB 1083 -----NVSLENFNIMADTKSOYIL-----NKKKNGTNNNTDYN 1117
QY 746 RPDGGEFTTSVAVOKFVODKPYGILYNHNDNEYLLODDWYILSKVENSPEYITFYVK 805
DB 1118 KE-----LKEHKKSNVYKDF-----AGKNTQF-----IK 1142
QY 806 GRNDAMGCGSVLYTSAVLPESILPELOTAQKVGRODNTEFIKTDNTCGPEPLVERL 865
DB 1143 KNELEFKYBOEV--TVLLNRYAVELEK-----NKFDTKN-----YSEGI 1181

```

```

QY 866 EKVVEGEERTIIEVEIEEVEKVRDKEVTLFSKLFEGFKELQNDENFRELKSEMD 925
DB 1182 IKELKDHANFTTQAKSEKKNMEINQDRI-----EDVAKNNKSNKA 1226
QY 926 VLD-GKMEATEVEKLEGRALPIKLMAYATHCF--TSPCHDIRFFSSDDIGRLGITR 982
DB 1227 ILDIOLSVPEFKI-----KFLKIKDLRTKSDCLKETKDIETIKISMLIDQEFKL-IEN 1280
QY 983 KRIINGFELKILPEIOSADLRTTGGSSRPLAFNRSGSKGFIPLVPLSKNELKELTAP 1042
DB 1281 KNILNT-LEKLESLKNO-----KKNED-----QKKEIDDEVNSK 1314
QY 1043 L-----DLKLVGLACAFLLVPSADAVDALKTACILKGRIPLACI 1085
DB 1315 IKNIESVNQHKKNYIEGIVEKINELAKA-----KQDQESQOKLI--IPTINL 1362
QY 1086 ANPACANVACLOTGNNRPD-ETECO-----IKGDL-----FENSV 1121
DB 1363 ISPEKANDLEGIDTNKILGKRYNTEMNNIYEETIKSYDILTHYLETYSKEPIYEQIKMR 1422
QY 1122 VDEFNECAVSRKCVPRKSDLGEPAPDPSVLYONF--NISDFNGK--YITSGINFTF 1176
DB 1423 ITNQNELLNINNVAKAKSYLDIDIANEPDRIVTFKKNLNDVNDKFTNEY--SKVNGKF 1480
QY 1177 DAFDCOLHEFHTEGDKLVGNISWRIKTLDSGFTFRSAVOKFVODPNOPVLYNHNDNEYL 1236
DB 1481 DNISINNVRKSTDENLILNLOTKEW--YANIVSKRY-----SYK 1522
QY 1237 HYODDWYILSKIEKPEDEYIVYVYGRNDAMDYGAGAVYTRSSVLPNSIIPLEKRAK 1296
DB 1523 YEAEENFTIPKLAN-----SLNIQIKSSSGIDLEKKNINI--ALPYLDSQK 1568
QY 1297 SIGRDEFTFIRTDNTCGPEPALVERIEKTEVEGERI-----IYKVEIEEVEEKE 1347
DB 1569 ----DLTFRI-----PSP-----EXTSEYTKNISQSYMTLIDLKRSGLQKKEQOA 1611
QY 1348 VEVKGTENTLQRLABGFENELKODEENFVRELSEKEHPELDEIKM--EASEVEKL 1401
DB 1612 LMLT--FENRLHDKVQATNELK--DTLSLDKNNKBOILNKVKYLLHKSNEINLKL 1662

```

```

RESULT 6
ID 049525 PRELIMINARY; PRT: 1365 AA.
AC 049525.
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE Lmp1.
GN Lmp1.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmatetaceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MH56;
RX MEDLINE: 95369882.
RA JENSEN L.T., LADEFOED S., BIRKELUND S., CHRISTIANSEN G.;
RT "Selection of Mycoplasma hominis pg21 deletion mutants by cultivation
in the presence of monoclonal antibody 552."
RL Infect. Immun. 63:3336-3347(1995).
DR EMBL: U21962; AAA81013.1; -
SQ SEQUENCE 1365 AA; 154982 MW; 23BE4182 CRC32;

```

Query Match 2.1%; Score 156; DB 2; Length 1365;
 Best Local Similarity 16.8%; Pred. No. 0.57;
 Matches 209; Conservative 187; Mismatches 420; Indels 430; Gaps 52;

```

QY 272 TRSAVQFPVODPDLGALYNHNDNEFLHYODWYILSSQIENKBDYLFYVYGRNDAMD 331
DB 282 TRNOIQEINT-----NKNNP--NYSE-----LISQILSKRD-----SKNSVTIS 319

```



```

QY 687 GDFPVDPVSVLVOKFDMKDFSGKWFITRGLNFPFAFCOLHEPHTENKLVNISMIR 746
Db 954 DDELLTAAVEKER-----HATENK-VKNITEEMA 982
QY 747 TPDGFFTSAAVOKFVODEKYPGILYNHNEYLLODDWYILSSVENSPEYIFVYK 806
Db 983 SQD-----ESIAKLTKREKKAQOEHAQOQLDLOAEEDKVTNTAKTRLEQOV----- 1031
QY 807 RNDAMDYGVGSVLYTRS-----AVLPESILP--ELQTAQKVR----- 844
Db 1031 -----DDESGSLSEOKRLMDLEVRKRLGDLKLAOSIMDLENKQSPKIKKKQFE 1085
QY 844 --DENTFIKTNTGPE-----PIVEREK-----KVEGERTIIVEEI 883
Db 1086 ISQFSLKIDESLQAOLOKLIKELQARLELEEEIEERSAKARAKVERQADLSRELEEI 1145
QY 884 EEEVER-----VRDKEVTLFSKLEFG 904
Db 1146 SERLEBAGGATAAQLEMMKKRREAFQKMRDLEESTLQHEATAALRRKQADVAELQEQ 1205
QY 905 FKELODEENFLBELSKEMDVLGKMEATEVEKLFGRALPIRKLAAVATCTSPCHD 964
Db 1206 IDNLRITOKLEKESEKMEI--DLSNMEAVAKAGN--LEKM-----CRT--LED 1253
QY 965 RIRFF--SSDDGIGRLGTRKRNCTFLKILPIQSADLRTTGRSSR-----P 1012
Db 1254 QLESEKASDENSRQLODMN-----AQRARLOTENGESTRQLEKEEALVSQ 1299
QY 1013 LSAFSSGSKGTFDVLPLPSKNEKELTAPLLKLVGLACFLVPSADAVDAKLTAC 1072
Db 1300 LTRGQAFQTOQIEDL--KRVHEE----- 1321
QY 1073 LKGRIRLAKCIANPACANAVAC--LOTGNRPDETECOIKG-----DLFE 1118
Db 1321 -----EVAKAKALAAVOSAHPCDCLDEQYEEQEAALQORMSKANSEVAQWRKAYE 1375
QY 1119 NSVDEFECAVSRKCVPRKSDLGEPAPDPVSVLVONFENISDNGKWTITSGINPFDA 1178
Db 1376 TDAIORTLEELSESKKILAQRLQDAE-----STEANSK--CASLEKTKOR 1419
QY 1179 FDCOLHEHTEGD--NKLGVNISMRIKTLDSGFFTSAAVOKFVODPNQGVLYNHNDNEYL 1236
Db 1420 LQSEVEDLMIDGERANALAAWLDRKQRFD-----KVLADMKOK--YEESQAE 1466
QY 1237 R-YODDWYILSSKIENKPEYIFVYRGRNDAMDYGAVVYTRSSVLPSIILPELEKAA 1295
Db 1467 EAAKEARSLSTELFKMKNSY-----EALDH-----LETLAKREN 1501
QY 1296 KSTGRDSTFIRNTNGCPPEPALVERIEKTVEEGERIIVKEVEIEEVEKEVEKGRTE 1355
Db 1502 KNLQOELISDSEQLGETGKSIHEIEKAKKTVESEKAIQJALAEAGETLHEHSKIIRVQ 1561
QY 1356 MTLPO-----RLAEGFNELKODENFVRELSEKMEFIDEIKM--EASEVK 1400
Db 1562 LELNQVASEIDRKLAEKDEKEQIKRNSQRYLDSMOSITLSEVSRNDALRVK 1615

```

```

RX MEDLINE; 98049343.
RA KLENN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERAVAGE A.R., GRAHAM D.E., KYRILDES N.C.,
RA REICHSMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODER A., ZHOU L.,
RA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KATNE B.P., SYKES S.M.,
RA SADOW P.M., D'AMBREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESSE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic,
RT sulphate-reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL; AE001032; AAB90211.1; -.
DR TIGR; AF1032; -.
DR PFAM; PF00470; RecF; 1.
KW Hypothetical protein.
SQ SEQUENCE 886 AA; 103633 MW; BF6BE847 CRC32;

Query Match          2.1%  Score 155.5;  DB 1;  Length 886;
Best Local Similarity 20.6%  Freq. No. 0.34;
Matches 137;  Conservative 85;  Mismatches 221;  Indels 221;  Gaps 27;

QY 797 EDYIFVYRGRNDAMDYGGSV--LYTRSAAVLPESILPELOTAQKVRDENTFIKTD-N 853
Db 152 EDY-----ENAMKNIGAVIRMLERKERKELTSEBQIKQKEK-----KAEIE 197
QY 854 TCGPEPLVERLEKKEVESEKRTI--KEVEIEEVEKVRDKEVTLFSKLEGEKELOR 910
Db 198 RISEEKSLSEIREKLSSEVRNLESRLKELEHKSRLSELSKROSSVLOEV-----R 249
QY 911 DEENFLRELSKEEMDV---LDGLKMEATEVEKLFGRALPIRKLAAVATCTSPCHDRIR 967
Db 250 GLEEKRLREKLEKVEVERIEDLEKKAKEVKEL-----KKAERY- 290
QY 968 PFSSDDGIGRLGTRKRNCTFLKILPIQSA--DLNTTGRSSRPLSAFRSGSGIF 1025
Db 290 -----SILEKLSLEINQALRDYKREGDLFREAGIQAOQKKAEE 329
QY 1026 DIVPLPSKNEKELTAPL-----LLKLVGLACFLVPSADAVDAKLTCTC 1070
Db 330 D-----NSKLEITIKRIEELRELERFEKSHRLE-----TLKPKMDRMQGIK-- 373
QY 1071 ACLKGRIRLAKCIANPACANAVACLOTGNRPDETECOIKGDLFENSVDDEFNCAV 1130
Db 373 -----AKLEENKLTLPDKYE--KMYDLISKAKEE--KEITE 404
QY 1131 SKKKVPRKSDLGEPAPDPVSVLVONFENISDNGKWTITSGINFPFAFCOLHEPTEG 1190
Db 405 KKLKLIARSSLSKTRGAOLKRAVE-----LKSABERTCPVCGRELDDEHRKN 451
QY 1191 -----DNKLVNISMRIKTLDSGFFTSAAVOKFVODPNQGVLYN-- 1231
Db 452 IMAEYTRMKRIAELEAKADIEKLEKLEKVEKALEKQETVLYKQOMDEKALANEL 511
QY 1231 --HDNEYLHYDDWYILSSKIENKPEYIFVYRGRNDAMDYGAVVYTRSSVLPSIIL 1288
Db 512 SSHDAE-----KLSAESE--YKXVERLDJLGRQOKILLSSA---SRI 550
QY 1289 PELKAAKSTGRDSTFIRNTNGCPPEPALVERIEKTVEEGERIIVKEVEIEEVEK 1348
Db 551 KELSSSLREI-----EAL--KNVESRGGLHKKIRREGEESIEELEEREV 593
QY 1349 EKVGRTEMTLFLQALAEFN--ELKODENFVRELSEKMEFIDEIKMSEASEVEKLFGR 1405
Db 594 -----QSLRPYNNKMLEKDAESRLSEFLKRR--KLDELSAIAKLEANGRA 641
QY 1406 LPIR 1409
Db 642 EIR 645

```

RESULT 9
 025662 PRELIMINARY: PRT: 1939 AA.
 ID 025662:
 AC 025662:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 07, Last annotation update)
 DE REPEAT ORGANELLAR PROTEIN.
 OS Plasmodium chabaudi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-96V.
 RA WERNER E.B., TAYLOR W.R., HOLDER A.A.:
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U43145; AAC63403.1;
 SO SEQUENCE 1939 AA; 229001 MW; 5FCAEAC3 CRC32;

Query Match 2.0%; Score 151; DB 5; Length 1939;
 Best Local Similarity 17.5%; Pred. No. 1.7;
 Matches 189; Conservative 167; Mismatches 299; Indels 422; Gaps 50;

QY 380 VERLEKTAEGEKLKEAVEIEEVEKEVERVDTMTLFORLLEGKELQD--EEN 436
 DB 95 VKEVEKIELEKEKEL--KLEKEKQINKKEKEKSEFIRKQMLLEKELNINKEN 151
 QY 437 FVEE-----LSKEKEILNELQWE--ATEVEKLFGRALPIRKULMALAPHSNLANHE 487
 DB 152 KINKKEIITILKREK--LNDISEYEIEKKKEKEL--NVEVTNIKMSL-----D 196
 QY 488 TIKYVYGSKLPGHKRPSWGMEDYFGSIYVAKICSSRRIRPKRSRITCCGDSRGLQLF 547
 DB 197 KLTCEVOEKKNLEKIN-----KKVIEKENNELKEEFKKEKKEIIESLDG----- 243
 QY 548 SHGKHNLSPHNSINQVVPKNSGCKFPKQVALVWMEKQGFATAIVAFILSVASKADA 607
 DB 243 -----TIN-----DKNAVEKLEI-----SFEKKRM 264
 QY 608 VDAKTKCTCLLKECRLELAKCISNPACANVACLQTCNNRPDETCQIKGDLFENSVD 667
 DB 265 IEMLD-----SKLIEKENFAN-----KQAKLE 287
 QY 668 ENNECAVSRKCYPRKSDVGFPPDPSPVYQKFDKQDSGKMFITRGILPTDAIDCQL 727
 DB 288 KENDEIIEKLKDIESRE-----KDFRSK-----EKFASMEDEL 321
 QY 728 HEFHEENKLVGNLS-WRIRPDGFFTRSAQVQFVODPKYPGILYNHNDYELLYODDWY 786
 DB 322 NTLKSDLSKNAQCOVEYKLEIKD--LSQSLVERE-----IPEIKNEY-----DD-- 366
 QY 787 ILSKVENSPEDYIFVYKGRANDAMDGYSVLYTRSAYLPESTIPELQTAQKGRDEN 846
 DB 366 ----KINMKKEKLSI-----ND--KGIDMTVLHSEEEKI-NKLKEKEKELNEIHKKNY 413
 QY 847 TFIPTDNTCGEPPLVERLEKKEVGEGERIIEKEVEIEEVEKVR-----DKEVT 896
 DB 414 LEID-----TIKNELEKEEVEKKKKAHYEVNLTLEIK 449
 QY 897 LFSKLFEKFLQDE-----ENFLRELKSEMDVLDGLKMEATEVEKLFGRALPIRKIMA 952
 DB 450 LLEKTEKADKGHKNELNELNOLSKLNKEK--DNINKNEMFL-----491
 QY 953 VAHCHFTSPCHDRIRFTSSDDGIGRLGITRKIRINGTFLTKLPIQADLRITGGRSSRP 1012
 DB 491 -----NDKISSLSNME-----VN-----ILKKDKQTLG-----513
 QY 1013 LSAPRSQSGKIPDIVLPSPKSNELKELTAPLLKIVGLACAFILVPSADVAIDLKTCAC 1072
 DB 513 -----NDIKTLN-----DLINLNKK--528

QY 1073 LKGRIEELAKCIANPACANVACLQTCNNRDE-----TECQIKGDLFENSVD 1123
 DB 528 -----EINTSDKNKMKMEDLAMEKEGKC-----VVID 558
 QY 1124 EF-----NECAVSRKCYPRK--SDLGEPAPDPSPVLYQNFNISPENKWTITSGINPIF 1176
 DB 559 EIEKKYKNEIFMLEEKLKEKENYADLND--EISLRNSIYKE--KEFI-----604
 QY 1177 DAFDQOHEFHTGENDKIKLVGNISWRIKTIDSGFTTSAQK--FYQDPNQGVLYNHDN 1233
 DB 604 -----EKKEFI-----ENKI-----NLFKNNEFEKNKIYENELNSLRKQDNQ 642
 QY 1234 EYLHYODDWYILSKKIENKEDYIFVYGRGRANDAMDGAGAVYTRSSVLPNSIPELEK 1293
 DB 643 GLIKQIDELNTQKLTREK--YLQLY-----NDN-----MHMFRSICIKIDMPSSEN 687
 QY 1294 AAKSIGDEST-PI--RTDNTCGEPALVRIKTYEGERIIVKVEEIEEVEKEVEK 1350
 DB 688 IKGSDLVDFVYATIKRDESSSDANPDTHK--EMVAELEKRAAIVALEEKHKEIK 745
 QY 1351 VGRTEMTLFORLLEGFNE--LKODEENFRELKSEMEFLDEIKMEASEVEKLF 1403
 DB 746 LGEGHKEVYLTLGQKHKEFTIILEKHKVDVYTKLGQHKENIILKEBEHKDVYTKLG 802

RESULT 10
 013246 PRELIMINARY: PRT: 3113 AA.
 ID 013246:
 AC 013246:
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE NUCLEAR PHOSPHOPROTEIN MITOSIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RC SEQUENCE FROM N.A.
 RC MEDLINE; 95379843.
 RA ZHU X., MANCINI M.A., CHANG K.H., LIU C.Y., CHEN C.F., SHAN B.,
 RA JONES D., YANG-FENG T.L., LEE W.H.:
 RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
 that is specifically involved in mitotic phase progression."
 RL Mol. Cell. Biol. 15:5017-5029(1995).
 DR EMBL: U30872; AAA82935.1;
 SO SEQUENCE 3113 AA; 357281 MW; 605779C3 CRC32;

Query Match 2.0%; Score 147; DB 4; Length 3113;
 Best Local Similarity 18.4%; Pred. No. 5.6;
 Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 IVERLEKTAEGEKLKEAVEIEEVE-----KEVE-KVBDTEMTLFORLLEGKELQ 432
 DB 946 LSETLSLEKKEKSSITSLNKEIEBELTOENGLKELINSLNQKMNLIKQ-SESFANYID 1004
 QY 433 DEENFVREL--KEKEIL-----450
 DB 1005 EREKSISELSDQYKQEKILILQRCBETGNAYEDLSQKYKAAQKNSKLECLINETSICE 1064
 QY 450 ---NELQMEATVEKLFGRALP--IRKULMALAPHSNLANHEPIKYYVGSKLPGH-----501
 DB 1065 NRKNLE-----QKKEAFKEHOEFITLKAFAERQONIMLELEYOOLRSBEMDNQNS 1120
 QY 501 KRFSGMEDYFGSIYVAVKICSSRRIPRYFRKSPRI-----CCGDSRGLDLSHGK 551
 DB 1121 KEGAGSLQKQINTELEQNKMKQKEVNDLQENQMLKWKTKHEQCNLESPIR-----1175
 QY 552 HNLSPHNSINQVVPKNSGCKFPKQVALVWMEKQGFATAIVAFILSVASKADAYDAL 611
 DB 1175 -----NSVKEKESERNQ--CNFKPQMDLEVKIISLDSYNAQDVOLEAALRNKELKLOE 1227
 QY 612 KTCITCL-----LKCEKLEL-AKCSINP-----ACAN 637

```

Db 1228 KEKECLOHLOTIRGDLSTNLDMOSQELSGIKDCEIDAEKYEYISGPHELSTSONDNH 1287
QY 638 VAC-LOTCNNRPDETE--COIKCGDLFENSVDDEFNECAVSRKCY--PRK--SDVGF- 690
Db 1288 IQCSLOTIMKRLNELBKICEIILQAEKYE--LVTEIND--SRSECTIATRKMAEYKGL 1342
QY 690 ----PVPDSVLYQKFDKMDFSKMFITRGLNP-----TFDADC 725
Db 1343 NEVKIINDSGILHGEVLDEIPGGEF--GEOPNECHPVSALPDESNSEYHLLTSDKEV 1399
QY 726 QLEHEFTEENKLYGNLSWRIRTPDGGFTFSRQVOPKYPGILYINDNYLLXOD-- 784
Db 1400 OMHFAELQEK-----FLSLQSEHKTLHQHOC--MSSKMSLQTYVDSL 1441
QY 784 --DWYLLSKVENSPEDIYFYVYKGRND-----AMDGYGGSVLYTRSAVL 826
Db 1442 KAEIVLSTNLRFQODLVKEMQIGLEGLVPSLSSSCVPSDSSLSIGDSSFY--RAL 1499
QY 827 PE-----SIPELOTAQ--KYGRDENTFIKTDNCGPEPL-----VERLEKVEEGE 873
Db 1500 EQTGDMSLISNEGAVSANOCSVDYFCSSIQDENLTKRETPSAPAKGYEELSLCEVYR 1559
QY 874 RTIIEVEEIEEVEK--VRDREVTLSKLFEG-----FKELORDEEFTLEISKE 922
Db 1560 OSL---EKLEEKMESOGINKKKEIQELQLLSERKQELDLRKQYLSENQMOCKLTSV 1615
QY 923 EMDVDGLKMEATEVEKLE-----GALPIRKIMAVAIH-----CFTS-- 961
Db 1616 TLEMESKLAERKQTEQJSLLEVARLOGLDLSRSRLGIDTBDALQGNESCDISKE 1675
QY 961 -----PCHDRIRFSSDDIGRLGITRKIRNGTFLKILTPFI-----QSADL-- 1003
Db 1676 HTSETTERPKHD-VHQJODKAOODLNDIKITETGALK--PTGECSEQSDPNIYE 1731
QY 1003 ----RTIGGRS-----SRPLSAFRSGSKGIFDIYPLPSK----- 1034
Db 1732 PEGEDKTQGSSECSISLSPSGFNALVPMDFLQNOEDIHNLQIRVETSNNMLRLHVED 1791
QY 1034 ----NELKELTAPILKLVLGACAPLIVSAAVVALKTCACLLGCKIELAKC 1084
Db 1792 RDRKVESLLENKELDKLHJQEVOLMT-----KTEAC-----TELEKI 1830
QY 1085 IANPACAAVAVCLOTCNNRPDETECOIKCGDLFENSVDDEFNECAVSRKCYPRKSDLG- 1144
Db 1831 V-----GEL-----KK--ENSDLSE 1843
QY 1144 --EPPADPVSUYONFNTSDNGKWYITSGINPTFDADCOLHEFHE--GDNKLGVN 1197
Db 1844 KLEIFSCHQHELLOREVETSE-----GLNS-----DLEMHADKSSREDIGDNVAKVN 1889
QY 1198 ISWRITLDSGFTFSRQVOPKYPGILYINDNYLLXOD--YQDDWYIISKLENK 1252
Db 1890 DSMKREFLD-----VENELSRIRSEKASIEHBALYLEDLEVOVTEKLCLEKNENK 1941
QY 1253 P-----EDYIFYVYRGNDAMDGYGGAVYTRSSVLPNSIPELEKANSIGHDFSTFI 1306
Db 1942 QKAVLCCLEELSVTSEKNOL--RGEIDTMSKKTITLDOISEKMEKETOELSHQSECL 1998
QY 1307 RTDTCGEPALVRIETKVEGERIIVKEVEIEE--VEKVEKVGTEMTLTPORLA 1363
Db 1999 HClVAEVEVEKTELLQTLSSDVSLLKDKTHIQEKLQSLSEKSOALSJLKCELENOIA 2058
QY 1364 EGENELKODEENVRE-----LSKEMEFLDEIK--MEASVEYK 1400
Db 2059 ----QLNKKEKELLVKESESIQARLSSESDYERKLNVSKALEAIVEX 2099

```

```

DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, last annotation update)
DE NMDLP.
GN NMDLP.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972.
RA BARRELL B., RAJANDREAM M.A., WALSH S.V.;
RU Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; 250046; CA30372.1; -.
DR SGD; L0001287; NMDL.
DR PFAM; PF00169; PH.1.
SQ SEQUENCE 2748 AA; 313030 MW; FEEETEDD CMC32;

Query Match          1.9%; Score 146; DB 3; Length 2748;
Best Local Similarity 17.9%; Pred. No. 5.4;
Matches 283; Conservative 232; Mismatches 581; Indels 486; Gaps 72;

QY 62 SSFSDSHCKDKSQISIDTSFEIQRFDKRGMLILEKQROF-----IQALIVLC 115
Db 748 SAVEDLVCKKENDV-----EFLKESAKLGHIVSSEYSELQRRYSELEKEVQPS 800
QY 116 TEYIVRVDVADALKTCACLLKECRTELAKCIANPSC-----AANVACLOTCNNRPDET 169
Db 801 LAVIVEHAKATPH--HLLSDAYEELVCKKENDPMEFLKESAKLGHIVSNAYSEL 856
QY 170 EEOI-----KCGD--LFENSVDQNECAVSRKCYPRKSDYGEFVPVDRNAV 215
Db 857 EKKLEQPSLAVIEHAKATDHLHLLSDAYEDLVK-----KENSUV-EF----- 900
QY 216 VONFNKDFSGKWIYITSGINPTFDADCOLHEFHENDKLVGNLFWRIKTLDDGFFTRSA 275
Db 900 ----LKEKSARKLGHIVSNAYSELEKLDQ-----PSLAVIEHAKATDHLHLLSDA 948
QY 276 VQTFVODPPLPALNHNHNEFLHYO-----DWYLLSSQIENKPPDYIFYVYRG 324
Db 949 YE-----DLVCKKENDPMEFLKESAKLGHIVSNAYSELEKLDQPSLEYIEHAKA 1002
QY 325 RND--AMDGYGGSV-----IYTRSPILPESIIPN-----LOKAASVGDRF-- 364
Db 1003 TNHLLSDSAYEDLVCKKENDPMEFLKESAKLGHIVSNAYSELEKLDQPSLEYIE 1062
QY 364 ----NFTITDNS-----CGPEPPLVERLEKTAEEGCKLLIKEAV-ELIEVEYE-- 409
Db 1063 HAKATNHLHLLSDSAYEELVCKKENDPMEFLKESAKLGHIVSNAYSELEKLDQPSLE 1122
QY 409 --VEKVPDTEMTLFGRLLEGKELQODENFVRELSEKEKILNE--LOMA-TEVKLFL 463
Db 1123 YLVEHAKATNHLHLLSD--SAYEELVCKKENDPMEFLKESAKLGHIVSNAYSELEK 1179
QY 464 GALPIRLRLMALAFNSFLANHETIKYVGSKLPGHRSFGWEDFGSIVAK----- 519
Db 1179 ----KLEQ--PSLAVIEH-----AKATDHLHLL--DSAYEDLVCKKENDPV 1217
QY 519 ----ICSSRIPIRYFRKSPRICCGIDSGCL-OLFSHGKHNLSPAHSINQNV 564
Db 1218 EFLKESAKLGHIVSNAYSELEK-----LEQPSLAVIEHAK--ATDHLHLLSDA 1268
QY 565 PKGNSGCKFPKDVALMVEKWKQGFAKTAIYAIPLTASVASKAD-----AVDAKLTQCL 617
Db 1269 YEDLVCKKENDPMEFLK-EKSAKLGHTVSNAYSELEKLDQPSLEYIEHAKATNHL 1327
QY 618 LKECKLE-LAKCISNPAC-----AANVACLOTCNNRPDETECOIKCGDLFENSVDDEFN 670

```

```

RESULT 11
ID 003767 PRELIMINARY; PRT; 2748 AA.
AC 003767;
DT 01-NOV-1996 (Tremblrel. 01, Created)

```

Db 1328 LSDSAVEDLVCKENPDMFLKESAKIGHTVSNKSELEKLEQPSL----- 1378
 QY 671 ECASRRKKCVPRK-SDVGF-----PVPDPSVLVQKFDKDESGKFFITRGILNPTIDAF 723
 Db 1378 EYLVAHQIOQSIKISIDFMTLANPSMEDMASKLQKLEYOIVSNDYI--ALNTKEKP 1435
 QY 724 DCQ-----LHFEFH-----TEENKLVGNISWRIRTRTDGCGFTTSAAVQKFPVD--PKXPGILYN 773
 Db 1436 DVELLRSLKGVHIIDTTYNELVSNFN-----SPTLKEIEEAKSKGRILI 1482
 QY 774 HDNEYL-----LYODDWYILSSK---YENSEP-----D 798
 Db 1483 EPNEVLDMNRATPTPSKEIDNFCQIGCYALDSKEYERLKNLSNPSSKFIENMALLD 1542
 QY 799 YIEF---YKGRNDAMDGSGSVLYTRSAVLPESIIPELOTAQKVGDEFTFKIDNTC 855
 Db 1543 LVLVDKTYOAMKD-----NASNKKSLIPSTKAL-----DFVTM----- 1577
 QY 856 GPEPLVRLKKEVGEGRITIKVEEIEEVEKVRDKEVTLFSKLFEGFELORD--- 912
 Db 1577 PAPOLASAKSSLOK-----RTLSDIENELKAL-----GYAIREKNLPN 1616
 QY 912 -EENFLRELSKEEM-----DVLGDKMEATEVEKLEFG-----RALPIR 948
 Db 1617 LEKPIVDNASKNDVNLGSKFSLVPLSTEEYDNMRKHTKILNIGDSIDELKECKEY 1676
 QY 949 KLMAVAITH-----CFTSPCHDIRFFSSDDGIRGLITRRIRINGITLLKILPIIOSAD 1001
 Db 1677 QMLIISKHDYERKOBATEINPEGEFLEKASALGY-----ELVSEVLDKRMKQMTDSD 1729
 QY 1002 LFTTGSRSLPSAFRSGSKGIFDIPLPSKNELKELAPLLEKLVGLACAFIIVPSA 1061
 Db 1730 IYMOEKARN-----EMVLL--RNEKE--ALQKKIEPSTIFILEKAA 1770
 QY 1062 -----DAVALKTCACILKGCRIELAKCIANPACANVACLOTCNNRPDETCQICG 1114
 Db 1771 GNMKILVQIEYDEL-----IRKC-----NHPTMELEESCH 1802
 QY 1115 DLFENSVNDEFNECAVSRKKCVPRK-SDGEPAPDPSVLQVONFISDENGKWIYITSGN 1173
 Db 1803 HL--NLVLLDQNESTLRPLENRVEDLINTLSKLNIAIPNTIYODIGAYE-----N 1855
 QY 1174 PTFDAFCQHEFHTEGDKLVGNISWRIKTLDGSGFTTSAAVQKVPDQNGVLY--- 1230
 Db 1856 PNFDIYAKDSLNR-----MDYVIAISQDYELWAKERQIDYLIKIS 1896
 QY 1230 -----NHQNEYLHYODDWYITLSKIEKNPEDIYFYIYGRNDAMD 1270
 Db 1897 SEKIDHIVPLSEYNLMVTNYPNPSLSYLKERAVLNHILIKEDDYKNIL-----AVSE 1950
 QY 1271 YGGAVVYITSSVLPNIIIELEKAASIGRDFSTFIRNTGCPALVERIEKTEVEE 1330
 Db 1951 HPTVHLSEKASILNKVLVDKD-----DRTMSRSIE---KPTDPLSTKALSMG- 1998
 QY 1331 RIIVKEV-----EIEEVEKEVEKVGRTMTLFORLAGFNLKODEENFVRLSKEM 1385
 Db 1998 KILVNSTHNRNKLSEPDSEFLIMAKAEGGLIISEKYSSELROQIDRPSLDVLKKA 2057
 QY 1386 EFLDEIKMEASEVEKLGKALP 1407
 Db 2058 AIFDSIIVENIEYOQLVNTTSP 2079
 RESULT 12
 Q26024 PRELIMINARY; PRT; 839 AA.
 AC Q26024;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE TOPOISOMERASE I.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

RN [1]
 SEQUENCE FROM N.A.
 RX MEDLINE; 9600126;
 RA TOSH K.T., KILBER B.J.K.:
 RT "The gene encoding topoisomerase I from the human malaria parasite
 RT Plasmodium falciparum".
 RL Gene 163:151-154(1995).
 DR EMBL; X83758; CAA58716.1; .
 DR PFAM; PF01028; Topoisomerase_I. 1.
 KW Isomerase; Topoisomerase.
 SQ SEQUENCE 839 AA; 98109 MW; 1CAFA3F0 CRC32;
 Query Match 1.9%; Score 144.5; DB 5; Length 839;
 Best Local Similarity 17.3%; Pred. No. 1.3;
 Matches 191; Conservative 176; Mismatches 390; Indels 347; Gaps 46;
 QY 352 LQRAASVGRDYNFTTTDSCGPEPLVERLEKTAEGSEKILLKAEVET--EEVEKEV 409
 Db 24 INKIKWLG---NN--KSCNSRSKESIKOKNSNSELGKIKNTKSLGKKEEKKKOI 78
 QY 410 EKYRDTMTLFCQLLGGFELQODEENFVRLSKKEKELNLEOMATEVEKLFGRALPI 469
 Db 79 SKRSNLEKKNLKEGKKRYEKKSRIV---KDTKLNTVIRKETONNK-----PK 128
 QY 470 KRLMALAPSNFLANHETIKYVGSKLPGHKRFSWGWEDYGSIVAKICSSRRIPRYF 529
 Db 129 KILKKS---EENF--EPINRW--WEKIDQDIDQNNY----- 159
 QY 530 KRSPRICGLDSGLOLFSHGKHNLSPAHSINONPKNSGCKFPKDVALLMWEKKGQTA 589
 Db 159 -----LEHRLG-IF-----SPPY-VQHHVP-----IFYRSIKIEL--NA 188
 QY 590 KITATFELLISVASKADAVDAKTCCLKECRLELAKCISNPACANVACLOTCNNRPD 649
 Db 189 KSEELATYWCs-----AIGSDYCTKEKFLNFKTFIN-----SLENDNIIO 231
 QY 650 ETECOIKCGDLFENSVNDEFNECAVSRKKCVPRKSDVGFVPDPSVLVQKEDM----- 704
 Db 232 ENETKLLKKGDLSNKFEID-----FWPIKDHLLKREELKNTKEE 271
 QY 704 KDSGKMFITRGLNPTDAFCQHEFHTEENKLVGNLSKRITPBG-----FFIRSAV 758
 Db 272 KEEKKRMKMERKLPYATVLDWIRKISSNKAEPGIFRGRGHPGOGILKKRIFFEDV 331
 QY 759 QKFVODPKYPIGLYN-----HDNE--YLLYODDWYITLSKVENSPEDIYFYIKG 806
 Db 332 INISKDAPVRPIYDNNCGHMGDIYHDNKVTLAYIKD-----SINDQIKTFIS 381
 QY 807 RNDAMDGYGGSVLYTRSAVLPESIIPELOTAQKVGGRDFNTFIKT---DNTGCEPPLV 862
 Db 382 AQSKFKGYKDMKEYNAR-----KLFSCVHKIRREDYKNNKNNIIDQOLGTAVYIL 433
 QY 863 ERLKVEEGRTTIKEVEIEEVEKVRDKEVTLFSKLFEGREKLODEENFVRLSKEM 922
 Db 434 DFLALRV--GGEK---DIDEADTV-----GCCSLVEHISFAHDIPFK 472
 QY 923 EMDVLQDKMEATEVEKLFGRALPIRKILMAVATHTCFS---PCHDRIRFPSSDGIGRLG 979
 Db 473 SVD-----SKQKTDKDEKYNKIPLPNTNLSISSECCITLDFLGKDSIRYNT----- 521
 QY 980 ITRKRINGTFLKILPIIOSADLRTTGGSSRPLSAFRSGFSKGIPIVLPKSNK--LKE 1038
 Db 521 --VKIDQAVYINII-----IFCKNNRNDGVPDQITCSKLNLEYKE 559
 QY 1039 LTAPLLKLVGLACAFIIVSADAVDAL--KTCACILKGCRIELAKCIANPACANVAC 1096
 Db 560 IMPTLSAVFRFVYNASTLDDQALKRIKEVIGKITYSLYSG-ETELNR-----SK 607
 QY 1097 LOTCNNRPDEFTLQIKCGDLFENSVNDEFNECAVSRK---KCVPRKSDGEPAPDPSV 1152
 Db 608 KRKSHLSDINILSDASDSTINDVNNEDYDENGINKRLSAITVYGRKNDVDKNSP-IEV 666

```

QY 1153 LVONENISDFNGKWTITSGINPTFAFCQOLHEPHEDGNKLVGNISMRKTLDSGFTTR 1212
DB 667 DVSNNI-----ELINFTNNAREVAIIICNORSPKCHDTTM 703
QY 1213 SAVQKFPVQDPNPGVLYNHD-NEYLHYQDDWYILSSKIENKPEDEIFYYGRANDAMGY 1271
DB 704 SKIKKOIE-----LYNEDIKEKKY-----LOHLKNSDKKFIIV-----739
QY 1272 GGAVVYTRSSVLPNSIIPLEKAASIGDSTFIRTDNTOCPPEPALVERLEKTYVEGGER 1331
DB 739 -----SKVSTLDGTLRP-----NKYKENMK 758
QY 1332 IIVKEVEELEEEVEKEVEKRTENTLFOFLAEGFNEIKODENFVRELKSEMEPTLDE- 1391
DB 759 -----EESCKKKL-----ITLIKVEILLNNQKAVDNDKTIATLGTSKINYDPR 802
QY 1391 -----IKMASVEKLFQKALPIR 1409
DB 803 ITVAFCKKEFPIEKVFNNSLRK 826

RESULT 13
051465 PRELIMINARY; PRT: 2166 AA.
AC 051465:
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE HYPOTHETICAL 254.2 KD PROTEIN.
GN BB0512.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31.
RX MEDLINE: 98065943.
RA FRASER C.M., CASJENS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIERA R., WHITE O., KETCHUM K.A., DODSON R., HICKEY E.K., GWINN M.,
RA DOUGHERTY B., TOMB J.-F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERLAUGE A.R., OUCKENBUSH J., SALZBERG S., HANSON M.,
RA VAN VOGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTERBACK T., MATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
DR EMBL: AE001153; AAC66876.1;
DR TIGR: BB0512;
KW Hypothetical protein.
SQ SEQUENCE 2166 AA; 254243 MW; F2899A0A CRC32;

```

Query Match 1.9% Score 144; DB 2; Length 2166;
 Best Local Similarity 16.8%; Pred. No. 5.1; Mismatches 266; Conservative 275; Indels 500; Gaps 71;

```

QY 279 FVQDPDLGALYNHNDNFIH-----YQDDWYILSSQIENKPPDDIFYYGRANDAMGYGG 334
DB 641 -----SQINTEDEPISLIQIQDKGIELSESVFNDSIHI-----QKALDMS 686
QY 335 ---SVYTRSPYL-----PEIIPNLOKAKSVGRDFNNFITDSCGPEPLVERLE- 385
DB 687 WKDELLALNKLSDIVYSSEELLSSATLKIESLEKDVN-----DRMEYV 730
QY 385 ---KTAEEGKELIKEAVE-----IEEVEKEVEKYRDEMILFQRLBLG 427
DB 731 LKTDGI-ESLYIEKKEKLDMSYSOSDAILGIEFIRQETLIKDSVPELDELNKKF 789
QY 428 KELQDEENFV-----RELKREKELINELQNEATE-VEKLGRLPIRLMA 475
DB 790 ---DKNNEFVSKIEEDYKLDKIESEEDILNFKSDLNFTIS-----KLDIV 836
QY 476 LAFHS-----NFLANHETIKYVGSKLGHKRFSGMEDIYGSIVY--AK 518
DB 837 SNKSNQKOIDPFLDRISKDILNKRDSINNEVDSKLS-----DWQSKLNETIVKLEN 889
QY 519 ICSSRRIP-----RYFRKSPRICGLDSRGLQFSHGK-----552
DB 890 LSSGKVDLDLDSEYTTIKELKFSIES-----LESYLEKIDEPNQAISDELQ 943
QY 552 ---HLSAHSINONVPGK-----NSGCKPEKDYALVWWEKWOQFAK-TAIVAFILSV 601
DB 944 DLMNHKRELEENLSKFAVAVLNSEEFYKEVDSILOKRTDIAFSQANIDITDLSL 1003
QY 602 ASKADAVDALKTCTCLKECRLELAKGISN-PACANAVAC-----641
DB 1004 NKFNDIN-----KEINKYNEVISTNYRGISENISKLENDIMHEILNLSRLTDR 1054
QY 641 -----LOTNNRPDETECOI-----KCGDLFE-----NSVDFEPECVSR-K 677
DB 1055 IDLSKGMENDLOKLESFVSKYQVEKFEKLVKDLDTODGEAKINKLVEIQYKKRSL 1114
QY 678 KCVPRKSDVGPDPVPVVLQVQEFMKDPSGKWPITRGLNTPFAFCQOLHEPHTEENKL 737
DB 1115 EADYRRIT-----DNDIM-----QAKERFEE-ITMELKNIESKSEFLNDLKEREFKL 1162
QY 738 V-GNLSWRIRIPDGGFTFRS-AVQKFPVQDPKPYGLIYNHNEVLLVQDDWYILSSVEN 794
DB 1163 IESNFEERST-----FLESGLAISK-IRDEITYKTLNSDEN-----LOTISE 1206
QY 795 SPEDIFYYYGRN-----DAMDGYGSAVLYTRSAVLPESTIPELOTAQKY 841
DB 1207 MDONFEIIEORSKDILEPEKELQDKIXCY-GFTNSOGEIKAGVBNKINKNHFDVCIKV 1265
QY 842 GRDFTFIKTNTGCPPEPLVERLE--KVEEGERTIIKEV-EEIEEVEKVRD---KE 894
DB 1266 ---NTLI-DDQYKVEIEIKRIDLSKLSIESTDSTIEKLNDSVSGCIDKIADFNFLKY 1320
QY 895 VTLESKLEFEGKELQDEENFLRELKSEEMDVLDGLMEATEVEKLGRLPIRLMAVA 954
DB 1321 IELEERCNEQOLNLENKIDNKIKAIIDNALSQYDGLKRYKADMDSESL-NSYATL 1378
QY 955 THCFSPCHDIRFSSDDGIGRLITRKIRNG-----TFLLKIILPIQ 998
DB 1379 SEERKSSKEMI--PELESQKLNKLNESDLNNVEKDVIRLKESYHNVSSHKLLEEDF 1436
QY 999 SADLTTGGRSRPLSAFRSGFSKGFIDVLPBKNLKELTAPLLKLYGLACAFIV 1058
DB 1437 FKDLKIRGEELKYLENFIASYNKIONLEYDLSK-----1473
QY 1059 PSADAVDALKTCAQLKGCRIELAKIANPACANAVACLOTNNRPDETECOIKCGDLFE 1118
DB 1473 -----LENKTELLIOSFRIDIPQKKDKDENFYDFTREFSKKMDQSELA---LME 1521
QY 1119 NSV---VDEFNECAVSRKKVPR-----KSDLGEFPAQDSVYLVOENFISDF-----NG 1164
DB 1522 TNLGKVDDEFVANNKOSIIDSFWFLNIKDVAKDWOEKSYSTIEKRINLAELIKSFEND 1581
QY 1165 KWTITSGINPTFAFCQOLHEPHTEGDN-----KLGVNISMRIKTLDSGFTT 1211

```

```

Db 1582 IFNKIGLESEFDEGFEIKAEIIFSNLONEAKKIEQSVHLDKFNIGE-SLNKLYLDEKFEV 1640
Qy 1212 RSAVOKFVODPNQ--PGVLYHNDNEYLYODD-----WYLLSSIEKKPE 1254
Db 1641 DFKLEKIDKXNKKTEDILIOAEVKFLTOCKLEDKIFELNOKLHEFTLLSSMLDYVR 1700
Qy 1255 DYIFVYVGRNDAMDYGAVVYTRSSV-----LPNSIPELEKAKSIGRD 1301
Db 1701 EMVDVSSDK-----ESFEGQELINKINISESEKISLYRNNEIETSLINENVSFSKISD 1756
Qy 1302 FSTIRIDNTGCPALVERIEKTEVEBERIIVKEVEIEEVEKEVKGRIEMTLFQ- 1361
Db 1757 LG-----LLEDELKSKLSKHS-----TSEIETIKSGLOEOIDK-----FEV 1791
Qy 1361 RLAEGFNLKODEENFVRELKSEEM-----EFLDEIK 1392
Db 1792 EFKKNHKLLEKVDNNILESKILNCDVOPNKFISEIK 1830

```

RESULT 14

```

096275 PRELIMINARY: PRT: 1558 AA.
AC 096275;
DT 01-MAY-1999 (Tremblrel, 10, Created)
DT 01-MAY-1999 (Tremblrel, 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel, 10, Last annotation update)
DE RESA-H3 ANTIGEN.
GN PF80915W.
OC Plasmodium falciparum.
OS Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RA GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
RA KOONIN E.V., SHALLOO S., MASON T., YU K., FUJII C., PEDERSON J.,
RA SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,
RA SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
RA FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum".
RL Science 282:1126-1132(1998).
DR EMBL; AE01424; AAC71972.1;
SQ SEQUENCE 1558 AA; 175658 MW; D49C2605 CRC32;

```

```

Query Match 1.9% Score 143.5; DB 5; Length 1558;
Best Local Similarity 17.98; Pred. No: 3.5;
Matches 206; Conservative 153; Mismatches 377; Indels 413; Gaps 50;
Qy 347 STIPMLQK-AAKSVGRDN-----NFTTNSGCGPEPIVE-----RLEKTAEG 390
Db 626 TILNVEETIASVTTFSTILEIEIOENTITNDT---EKELEHENVLSALENTQSEE 682
Qy 391 EKLILKEAVEIEEVEKEV-----EKVADTEMTLEFORLLEGFKLQOD-----EENFVREL 441
Db 683 EK---KEVIDVIEEYKEEVAITLIEVQAEBSASTITEIPENLENAVENSEVNAENL 739
Qy 442 SKEEELINELOMEATEVEKLFGRALPIRLKLMALAPSNFLANETTKIYVGSKLPGHK 501
Db 740 EKLINTVENTVADKVEEVEISEGSELENEMDKAF-----FSEIPDNVKGIOENLTG-- 793
Qy 502 RFSWGMEDFGSIYVAKICSSRRIPRYFRKSPRICCGDLSRGLQFSGH---KHLSPA 557
Db 793 -----MFRSIEIETIYQSEE--KVDLENVVSIIIDN--IEMKKGKLNKLENIIST 840
Qy 558 HSINONVPGK-NSGCKFKPDVALMWKMGQFAKTAIYAIFLLSVASKADAVDAKLTCTC 616
Db 841 EGVQETVTEHVQNYVYVDVDPAMK---DQF-----LGLINAGG----- 878
Qy 617 LKECRLEIAKICISNPACANACIQTCCNNRPDETCQ---IKGDDLFENSVVDEFNCA 673
Db 878 -LKEMEFNLVDVFKS---ESDVIITVEIKDEPVOKEVEKETVSIIEEMENIVDLEE-- 932

```

```

Qy 674 VSRKKCVPRKSCVGD--FPVDPVSVLVQFPMKDFSGKWFITRGILNPTFADFQQLHEFH 731
Db 932 -----EKEI-LTRKMLDAVEESIEISS-DSKEET-----ESI 961
Qy 732 TEENKLVGLSLRKIRTPDGGFTTBSAVOKFVODPYPPIIYNHNDNEYLYODDWYIISK 791
Db 962 KQEKQVSLVLEVDND---MDESVEKYLE-----LKNMEBELKDAVEINDITSK 1010
Qy 792 -VENSPEDEYIIVYKGRNDAMDYGGSVLYTRSAVLPESI--IPELQTAQKVGDFNTF 848
Db 1011 LIEETQE-----LNEVEALINDMEKLEKLEKA---LSDSKEI 1046
Qy 849 IKTDNTGGEPPVLELEKVEEGE--RTIIEVEEI-----EEVEKYVD-----K 893
Db 1047 IDAKD-----DTLEKVEIEEHDIITTTLDEVEVELKQVEBDKIEKYSDLKLEEDILK 1097
Qy 894 EV-----TLFSKLFEGKELQRDENFLRLSKSEMDVDLGLKMEATEVEKLFGRAL----- 946
Db 1098 EYKEIKELSEELIEDYKELKTIETDILEEKKEIEKDHEKEFEAEELIKOLEADILKEVS 1157
Qy 946 -----PIRKLMVAATGCTSPCHDIRFRFSSDDGIGRLGTRKRIINGTFLKIL 994
Db 1158 SLEVEEKKLEVEYHELKEVEHIIISGDH-----IKG----- 1190
Qy 995 PPIQSADLRITGGSSRPLSAFRSGFSKGFIDIVLPKSKNELKELTAPLLKLTVGLACA 1054
Db 1190 --LEEDDLEVDQDKSILDMLKGMELGDM-----KESLEDTAKL----- 1231
Qy 1055 FLIVPSADAVDAKLTACILCKGRIFLANCIANPACANVACLOTGN--NRPDTECOIK 1112
Db 1231 -----GERVESLKDVLVSSALGMDEQMKTRKAQRKLEEVILK 1269
Qy 1113 CGDLFENSVDYJNECAVSRKKCVPRKSDLGEPAPDPVSVLVONTNISDPNGKWITSGL 1172
Db 1270 -----EEYKEPKKKTITKK-----VRDIDK-----K 1292
Qy 1173 NPTFADFQQLHEFHTEGDNKLIVGNISWRKITLDGFTTBSAVOKFVODPNQGVLYNHND 1232
Db 1293 EPKDEIIVEYEMKDEDIDED-----IEEDVEED----- 1320
Qy 1233 NEYLHYODDWYIISKIENKPEYITVYVYGRNDAMDYGAVVYTRSSVLPNSIPELE 1292
Db 1320 -----IEEDKVEDIDEDID-----E 1334
Qy 1293 KAKSIGRD---FSTFIRIDNTGCPALVERIEKTEVEBERIIVKEVE-----IE 1341
Db 1335 DIEDIDGEDKDEVIDILVQKEKRIEYKKEKKLEKKEVSGKLGKHVDEPMKYQKID 1394
Qy 1342 EYEVEKEVYV--GRIEMTLFORLAEGFNLKODE-----NFVRELKSEMEFLDEIKM 1393
Db 1395 KEVDKESKALKESKNVT-----NVLKQNDQFSSVKKNFVKYVAFPIISAVAA 1445
Qy 1394 EASEVEKLF 1402
Db 1446 FASYVVGFF 1454

```

RESULT 15

```

092LV0 PRELIMINARY: PRT: 1819 AA.
AC 092LV0;
DT 01-MAY-1999 (Tremblrel, 10, Created)
DT 01-MAY-1999 (Tremblrel, 10, Last sequence update)
DT 01-MAY-1999 (Tremblrel, 10, Last annotation update)
DE CAG ISLAND PROTEIN.
OS Helicobacter pylori J99.
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RA ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMBL G.,

```

RA TUMINO P.J., CAROSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
 RA GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
 RA TRUST T.J.,
 RT "Genomic Sequence Comparison of Two Unrelated Isolates of the Human
 RT Gastric Pathogen Helicobacter pylori,"
 RL Nature 397:176-180(1999).
 DR EMBL: AE01481; AAD06047.1;
 SO SEQUENCE 1819 AA; 207505 MW; 88095345 CRC32;

Query Match 1.98; Score 142.5; DB 2; Length 1819;
 Best Local Similarity 19.0%; Pred. No. 4.9;
 Matches 267; Conservative 199; Mismatches 524; Indels 417; Gaps 68;

QY 178 LEFNSVVDQFNECAVARKKCVPRKSDVGFPPVDRNAVQNNMKDFSGKWTTSGLNPT 237
 DB 244 LFSRSIFHTF-----VPLEDKSSRFS-KDRNLVYND----- 274
 QY 238 FDAFDQLHEFHMHENDKLIVGNLTWRIKTLDGGEFTRSVQTFVODPDLPGALYNHNDNEFL 297
 DB 274 ----ELQIRQ---EYRNLKERNEKGNMIDKNLF-----ENDDPN--RTLYN----- 312
 QY 298 HYDDWYILSSQJENKPDYIIFYVYGRNDAMDYGGSVYTRSPITLPESTIPNLOKAK 357
 DB 312 -----YLNIAETEDK--NPLRAFYECISN-----GGN--YECCIKLIRD--KTIQDQMK 354
 QY 358 SVGRDNFNTTDSGCEPEPLVERLEKTAEEGK--LLIKEAVEL-----EEVY 405
 DB 355 KITEAVNDCL--KNAKTEEERIKCDJIKDENLKSLSLNOQKVOYALDCLKNAKTDEER 411
 QY 406 EKEVEKVRDTEM-TLFORLLEGEFKELQDENEVEFRELSEKEE-EILNELQMEATEVEKLF 463
 DB 412 KECTKLINDPEIEKREKRELEQKELQEKYKDCIKNAKTAENKCEKLSKKA--IERK 469
 QY 464 GRALPFRKLMLAPHSNLANHETIKIYVGSKLPGHKKFSGWEDYFGSIVANICSSR 523
 DB 470 QOALDCLKNAKTDEERKELKN--IPDILQKELLADM-----SVAKYKDCVSR 515
 QY 524 RIPRYERK-----SPRICGSLDSRGLOFSGKHNLSPASHINONVPR-----GNSG 570
 DB 516 ARNEKKEQCEKILLPEAKKLLLENALDCLKNAKTDEERKEL--KNLPDOLDSILAKES 574
 QY 571 CKEPKDVA-----LWWEKMGQFAKTAIVAIFLSVASKADAVALKTCTC 616
 DB 575 LKAYKDCASQAKTEAEKKECEKILLPE-----AKKLEEEAKESVAKYLDVCSQAKT--EA 628
 QY 617 LKEC-----RLELAKCISNPACANACLOTNNRPDETECO---IKCGDLFE 662
 DB 629 EKKECEKILLPEAKKLEAK---KSVRAYLDCVSKAKNEAEKKECEKILLPEAKKLE 684
 QY 663 NSVYDEFNECAVARKKCVPRKSDVGFPPVDRNAVQNNMKDFSGKWTTSGLNPTEDA 722
 DB 685 NQALD-----CLNNAKTDEERKELCLDLP-----KDLQKVLAKESVAYLDC 727
 QY 723 FDCQLHEFTEENKLVGNLSWRIKTPDGGEFTRSVQTFVODPDKYPGILY----- 773
 DB 728 VSKAKNEAEKKECE-----KLTPPE-----ARKLLEPAKSVAKYKDCVLRARNE 772
 QY 773 --NHNEVYLLYODDWYILSSKVENSPEDYIFYVYGRNDAMDYGGSVYTRSAVLPESTI 830
 DB 773 KKKQCEKILLPEAKKLEES--KKSVAKYLDCVSKAKNEAEKKECEKILLPEAKKLEEA 831
 QY 831 IPELOTAQKVGDEFTFKTNTCGPEPLVERLEKKEVEGEPT-----IKVEVEI 883
 DB 832 KESVAKAYKDCVLRARNE--KEKQEC--EKLLTPPEAKKLEESKSVAKYLDVCSAKKNEA 887
 QY 884 F-EVEKVAQDKVYLFESKLFEGFKEL-----QDENENFLRELSEK-----EM 924
 DB 888 EKKECEKILLPEA---RKLLLEAKESVAKYKDCVLRARNEKKECEKILLTPPEAKKLEEN 944
 QY 925 DVLQGIKMEATEVE--KLFORALP---IKRLMAVATHCTSPCHDKIRFFSSDDGIGRLGI 980
 DB 945 QALDCLKNAKTAEKRCVKVCLPKDLQKKVLA-----KESVAYLDCVSKAKKNEA 994

QY 981 TRKRINTEFLIKTIPQISADLTRTGGSSRPISAFRSGFSKGFIDVLPKSNELKELT 1040
 DB 995 ERKECE-----KLTP-----EARKLLEAKESVAKYKDCVLRARNE--KEKQCEKILL 1041
 QY 1041 APLLKITY-----GVLA-----CAFLVPSA-----DAVDALKTG-- 1071
 DB 1042 TPEARKLLEQEVKSVAYAYLDCVSRARNEKKECEKILLPEARKLLENQALDCLKNAKT 1101
 QY 1071 ACLKGRIEL-----AKCIANPACANACLOTNNRPDETECO---IKCGDLFENS- 1121
 DB 1102 EAEKRCVQDLPKLOKKVLAKESVAKYLDVCSVRARNEKKECEKILLPEARKLLEESK 1161
 QY 1121 -VDFEFCVASKKCVPRKSDVGFPPVDRNAVQNNMKDFSGKWTTSGLNPTEDA 1179
 DB 1162 KSVAYLDC--VSKAKNEAEKKECEKILLPEARKLLE-----AKESVKAY 1205
 QY 1180 -DCQLHEFTEGDMKLVGNLSWRIKTLDSGFEFTRSVQTFVODPQGVLYNHNDNYLHY 1238
 DB 1206 KDC-----VSRARNEKQ-----ECEKILLT 1226
 QY 1239 QDDWYILSSKIEKPKEDYIFYVYGRNDAMDYGGAVVYTRSSVLPNSIIP-----LEKA 1294
 DB 1227 PEARKLLEQEVKSVAKYLDVCSVRARNE-----KEKQCEKILLPEARKLLEKQ 1275
 QY 1295 AKSIGRPFSTFIRDTNCGPEPALVERIEKTYVEGRITVKEVEIEEVEKEVEKVGRT 1354
 DB 1276 RQOKRAIKDCLR--NADPNDRAAIMKCLDGLSDEKEL--KYLOEAREKAVLDCLKTART 1331
 QY 1355 E-----MTLFORLLEGEF--NELQDENEVEFRELSEKEE-----MEFLDE----- 1391
 DB 1332 DEERKQNTYSDLQEIQNKKAQNN--QLSKTERLHQSCLINDPPTDQDEALIQ 1388
 QY 1391 -----IKMEASEVEKLF 1402
 DB 1389 CLEGLSDSERALILGIRQADEVDRIY 1415

Search completed: October 13, 1999, 22:45:24
 Job time: 742 sec

us-09-075-375-1.rni

D b	1447	AAGATTGTGACRRR	1386
OY	1405	gagaagtgttgataaaagacgtctgatgatatgaagaaggltggaaaagagttggag	1454
D b	1387	RR	1358
OY	1465	aagttagagatcctggagacttgcctaagsttgtcgaaagtccttaagaagtlg	1524
D b	1337	RR	1268
OY	1525	caacaagtgaagaattcttgaggaattgattaagaagagaagaaattctgcat	1584
D b	1267	RR	1208
OY	1585	gaactccaatlggaagcgaactgtaagttaaagcttctggcgcgtaccagatlbg	1644
D b	1207	RR	1148
OY	1645	aaacttgataaatcttcgatgatctcacagaacatatatatagtcataatgatatgla	1704
D b	1147	RR	1088
OY	1705	galactagagaaacccaaaaa	1726
D b	1087	:: : ::::: RRRRRRRRRRRRRRRRRRRA	1066

US-07-915-246-1
; Sequence 1, Application US/07915246
; Patent No. 5401836

APPLICANT: Baszczynski, Chris L.
 APPLICANT: Fallis, Lynne
 APPLICANT: Bellmare, Guy
 APPLICANT: Bolvin, Rodolphe
 TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
 TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: McAndrews, Held, and Malloy
 STREET: 500 W. Madison St. 34th Floor
 CITY: Chicago
 STATE: Illinois
 COUNTRY: USA
 ZIP: 60661
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/915,246
 FILING DATE: 19920716
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Pochopien, Donald J.
 REGISTRATION NUMBER: 32,167
 REFERENCE/DOCKET NUMBER: 91 P 1125
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312 707-8889
 TELEFAX: 312 707-9155
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1505 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Brassica napus

Query Match	2.1%;	Score 41.2;	DB 1;	Length 1505;
Best Local Similarity	47.3%;	Pred. No. 0.084;		
Matches 124;	Conservative	0;	Mismatches 138;	Indels 0;
			Gaps	0;

US-08-323-170B-1
: Sequence 1 Ar

Sequence 1, Application US/063231/05
Patent No. 5733772

GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
TITLE OF INVENTION: Cloning and Expression of Plasmidum
TITLE OF INVENTION: Filicliparum Transmission-Blocking Target Antigen, Pts300
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993

ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

; INFORMATION FOR SEQ ID NO: 3:

INFORMATION FOR SEX ID NO: 3

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/471,033
3 FILING DATE:
4 CLASSIFICATION: 530
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: US 08/314,594
7 FILING DATE: 09-SEP-1994
8 PRIOR APPLICATION DATA:
9 APPLICATION NUMBER: US 08/218,018
10 FILING DATE: 23-MAR-1994
11 PRIOR APPLICATION DATA:
12 APPLICATION NUMBER: US 08/037,057
13 FILING DATE: 25-MAR-1993
14 ATTORNEY/AGENT INFORMATION:
15 NAME: Pace, Gary M.
16 REGISTRATION NUMBER: P-40,403
17 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: 919-541-8582
20 TELEFAX: 919-541-8669
21 INFORMATION FOR SEQ ID NO: 1:
22 SEQUENCE CHARACTERISTICS.

```

? MOLECULE TYPE: DNA (genomic)
? ORIGINAL SOURCE:
? ORGANISM: Bacillus cereus
? STRAIN: AB78
? INDIVIDUAL ISOLATE: NRRL B-21058
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1082..2467
? OTHER INFORMATION: /product= "VIP2A(a) "
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: 2475..5126
? OTHER INFORMATION: /note= "Coding sequence for the 100
? OTHER INFORMATION: kd VIP1(a) protein. This coding sequence is repeated in
? OTHER INFORMATION: NO:4 and translated separately."
? US-08-471-033-1

```

	Query Match	Similarity	Score	DB	Length
Best Local	Similarity	44.8%	Pred. No. 1.4		
Matches	143	Conservative	0	Mismatches	176
				Indels	0
				Gaps	0
QY	334	cctctaccaacaatacatgatgaaattcgatccacaatgatatttaattcttc	393		
Db	938	cccccttcttattatgaagcctgttaaatatagagaagatgcttttttcacctcc	819		
QY	394	cggattttaagatcttaagaagaagtccttcacagatctagccatlycaagataaa	453		
Db	878	cgtttgttttttttcttataatttaacttaacctaaagatagacatccgttatggaaa	819		
QY	454	tcctcagatctgaagatcgatatacaagttttgaagaataacaagaatttgatccaaag	513		
Db	818	tatatataattttatgaaaaccttcatagttgttttcattttcagtcagtggaacaaata	759		
QY	514	ggcattgactttgattcttgaaagaacatgagacaattctacatggtgctatcgattg	573		
Db	758	aaactggcctgtgaaggaatcaaggttagcgttaattcaaacagattattttattatgcattcc	699		
QY	574	gtttgcacatttgatcatcgtcccaagattgatgcgltgatgctcttaaacatttgct	633		
Db	698	caaaaacgacccgtttatggacactgtgtttttttaaacaagaagtttaaaaactcatcca	639		
QY	634	tgttactcaagaatgca	652		
Db	638	tattgagtacccaattcca	620		

US-08-471-044-1

ATTORNEY/AGENT INFORMATION:

```

Query Match          1.9% ; Score 37.4 ; DB 3 ; Length 6049;
Best Local Similarity 44.8% ; Pred. No. 1.4;
Matches 143; Conservative 0; Mismatches 176; Indels 0; Gaps 0:

QY 334 cctctaccaacetaatcatgatgaaattcgatccacaatggaatttaattcttc 393
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 938 CCCCCCTTTTATTAAAGAGCCGTGAATAATATAGAGAAGATGCTTTTTCACCTCC 879
QY 394 cggcttcttaacattcataagaacagttctcttcagattcagccatigcaagataa 453
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 878 CGTTGTTGTTTTTTCTTATTAATTTAACTATAACCTAAGATAGAGCATCCGTTATGGA 819
QY 454 tctcagatattgacatcatgatacaagtttbgagaaatacaagaatttgctcaaaagg 513
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 818 TATATTAATTTTATGAACACCTTCATGTTGTTTTCATTAATTCAGTCAGTGTGAAGTCAATA 759
QY 514 ggcagatcttgattctcttgaaagcaatgagagcaattcaataattgtagctgattg 573
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 758 AAATCGGCGGTGAAGCATATACAGGTTTAGGTTATCAACACGATTATTTATTAATGTCATCC 659
QY 574 gtttcagattgtatcgttcccaagatgtagccgltgatgctctaaactgtgct 633
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 698 CAAAAGACCGCTTTATGGCAGCTTGTTGTTTTTTTAACAAAAGCGTTAAAAACTCATCA 639
QY 634 tgtttactcaagaaatgca 652
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 TATTGAGTACCAAAATCCA 620

RESULT 12
US-08-463-483A-1/c
; Sequence 1, Application US/08463483A
; Patent No. 5849870
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziele, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Koslichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,483A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:

```

```

NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6049 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Bacillus cereus
STRAIN: AB78
INDIVIDUAL ISOLATE: NRRL B-21058
FEATURE:
NAME/KEY: CDS
LOCATION: 1082..2467
OTHER INFORMATION: /product= "VIP2A(a)"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2475..5126
OTHER INFORMATION: /note= "Coding sequence for the 100
OTHER INFORMATION: kd VIP1A(a) protein. This coding sequence is repeated in SEQ
US-08-463-483A-1

```

```

Query Match          1.9%; Score 37.4; DB 3; Length 6049;
Best Local Similarity 44.8%; Pred. No. 1.4;
Matches 143; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 334 cctctccacacatcaatgataaattcgatccacatgatatcttaattcttc 393
DB 938 CCCCCCTTTATTAAAGAGCGCTTAAATATTATGAGAGAGATGCTTTTTCACCTCC 879
QY 394 CGGTTGTTACATCTTAAGACAAGTCTCTCAGATCTCAGCATGCAAGGATAA 453
DB 878 CGTTGTTTCTTATTAATTAATTAACCTAAGAGATAGAGCATCCGTTTGGAAA 819
QY 454 tctcagatagcgcacatcgatcagaagtttgaggaataacaaagatttgctcaaaag 513
DB 818 TATATAATTATTAGAAACCTTATGTTTTCATATTCAGTCAAGTGAAGTACAATA 759
QY 514 ggcagacttgatcttgaagaagcaatgagacaattcacaattcagctcgtatg 573
DB 758 AAAGTGGCTGTAGAGATACAGTTAGTTATCAACAGATTATTATATGTCATCC 699
QY 574 gttgcacattgtatcgttccagaagttgatgcggttgatgctcttaaaattgct 633
DB 698 CAAAAACGACGTTTATGAGCATGTTGTTTAAACAAAGCGTTAAACATCTCA 639
QY 634 tgttactcaagaatgca 652
DB 638 TATTGAGTACCAAAATCCA 620

```

RESULT 13
US-08-471-046A-1/c

```

; Sequence 1, Application US/08471046A
; Patent No. 5866326
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziele, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J

```

```

TITLE OF INVENTION: Method for Isolating Vegetative Insecticidal
TITLE OF INVENTION: Protein Genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 5866326artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6049 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Bacillus cereus
STRAIN: AB78
INDIVIDUAL ISOLATE: NRRL B-21058
FEATURE:
NAME/KEY: CDS
LOCATION: 1082..2467
OTHER INFORMATION: /product= "VIP2A(a)"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 2475..5126
OTHER INFORMATION: /note= "Coding sequence for the 100
OTHER INFORMATION: kd VIP1A(a) protein. This coding sequence is repeated in
OTHER INFORMATION: NO.4 and translated separately."
US-08-471-046A-1

```

```

Query Match          1.9%; Score 37.4; DB 3; Length 6049;
Best Local Similarity 44.8%; Pred. No. 1.4;
Matches 143; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

```

```

QY 334 cctctccacacatcaatgataaattcgatccacatgatatcttaattcttc 393
DB 938 CCCCCCTTTATTAAAGAGCGCTTAAATATTATGAGAGAGATGCTTTTTCACCTCC 879
QY 394 CGGTTGTTACATCTTAAGACAAGTCTCTCAGATCTCAGCATGCAAGGATAA 453
DB 878 CGTTGTTTCTTATTAATTAATTAACCTAAGAGATAGAGCATCCGTTATGAAAA 819

```


TYPE: nucleic acid

```

;
OPERATING SYSTEM: PC-DOS/MS-DOS

```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ. ID NO.: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20
```

```
Query Match 1.9%; Score 36.8; DB 3; Length 32207;
Best Local Similarity 48.1%; Pred. No. 3.7;
Matches 104; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1359 tgaagccctcattgtggaaggttgagaacacagcggaagggcgagaagtgttgat 1418
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 19762 TGAGCAGCAGCAGCAGCAGATGAACAGCAGCAGCAGAGAGAGAGAGCAGCAGAGGAGCA 19703

QY 1419 aagaagagctgtagagatagaagaaggttgaaaaaagaggtgagaaggttagaatatc 1478
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 19702 GGAGCAGAGAGTTAGAGAGAGCAGAGCAGAGTTAGAGAGATCAGAGACAGAGAGTTAGAGA 19643

QY 1479 tgaatgaacttgttcagaaggttgcttgaaggggttaaggaggtgcaacaagatgaaga 1538
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 19642 GCAGGAGCAGAGAGTTAGAGAGAGCAGAGCAGAGTTAGAGAGCAGAGCAGAGAGTTAGA 19583

QY 1539 gaatttgtgaggaagcttgagtaagaagaagaaga 1574
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 19582 GGAGCAGAGCAGAGAGTTAGAGAGAGCAGAGAGGA 19547
```

Search completed: October 14, 1999, 00:42:50
Job time: 3426 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Comphen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 1999, 00:34:29 : Search time 24.66 Seconds
(without alignments)
654.388 Million cell updates/sec

Title: US-09-075-375-2
Perfect score: 7495
Sequence: 1 MALSHVFLCKEALNLYA.....MEASEVKLGKALPIRRVR 1412

Scoring table: BLOSUM62

Searched: 119832 seqs, 11428610 residues

Database: Issued_Patents_AA:*
1: /cgm2_6/prodata/1/iaa/3A_COMB.pep:*
2: /cgm2_6/prodata/1/iaa/3B_COMB.pep:*
3: /cgm2_6/prodata/1/iaa/PCTUS9_COMB.pep:*
4: /cgm2_6/prodata/1/iaa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	147	2.0	2482	1	US-08-328-254-6
2	135.5	1.8	1312	2	US-08-592-126-148
3	132	1.8	776	1	US-08-021-601-2
4	132	1.8	776	1	US-08-082-849B-2
5	132	1.8	776	3	PCT-US94-01624-2
6	124	1.7	1581	2	US-08-404-531B-28
7	124	1.7	1498	2	US-08-404-531B-28
8	115.5	1.5	1130	2	US-08-460-309-2
9	115.5	1.5	1130	2	US-08-460-309-2
10	115.5	1.5	1130	2	US-08-125-077-2
11	115.5	1.5	1130	2	US-08-125-077-2
12	115	1.5	1435	1	US-08-568-459A-4
13	112.5	1.5	3248	1	US-08-353-700-1
14	112.5	1.5	3248	3	PCT-US95-16216-1
15	111.5	1.5	3418	2	US-08-603-753D-4
16	109	1.5	3135	1	US-08-323-170B-2
17	109	1.5	425	2	US-08-986-963-2
18	108.5	1.4	3418	2	US-08-639-501-2
19	104.5	1.4	984	1	US-08-242-932-2
20	104.5	1.4	984	2	US-08-714-481-2
21	104.5	1.4	984	3	PCT-US95-06111-2
22	103.5	1.4	837	2	US-08-426-627-23
23	103.5	1.4	450	2	US-08-712-939B-1
24	102	1.4	656	2	US-08-441-139-15
25	101.5	1.4	650	1	US-08-325-071-56
26	101.5	1.4	688	1	US-08-325-071-57
27	101.5	1.4	984	2	US-08-257-073-3
28	101.5	1.4	984	2	US-08-184-009-120
29	101.5	1.4	3075	2	US-08-460-309-5
30	101.5	1.4	3075	2	US-08-125-077-5
31	101.5	1.4	984	2	US-08-458-356-120
32	101.5	1.4	392	2	US-08-904-031-1
33	100.5	1.3	528	2	US-08-160-670A-5
34	100.5	1.3	789	2	US-08-471-033-32
35	100.5	1.3	789	2	US-08-471-044-32
36	100.5	1.3	789	2	US-08-463-483A-32
37	100.5	1.3	789	2	US-08-471-046A-32
38	100.5	1.3	789	2	US-08-470-566B-32
39	100.5	1.3	789	2	US-08-838-219B-4

40	100	1.3	1196	1	US-08-144-121-4	Sequence 4, Appl1
41	100	1.3	836	2	US-08-426-627-6	Sequence 6, Appl1
42	100	1.3	1196	2	US-08-735-893-4	Sequence 4, Appl1
43	99.5	1.3	1209	3	PCT-US95-04589-107	Sequence 107, App
44	99	1.3	654	2	US-08-441-139-11	Sequence 11, Appl
45	98.5	1.3	1107	1	US-08-366-577-2	Sequence 2, Appl1

ALIGNMENTS

```

RESULT 1
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
;
GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patonlin Release #1.0, Version #1.25
;
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 08/141,239
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-328-254-6
;
Query Match 2.0%; Score 147; DB 1; Length 2482;
Best Local Similarity 18.4%; Pred. No. 0.0034;
Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

```

Thu Oct 14 07:57:25 1999

us-09-075-375-2.rai

Page 2

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

733 EENKLVGNLSWRIRTPDGGFFTRSAVQKFYDDPKYIPGILYNHD---NEYLLQDDMW-I 78
| | : | | :: : | | : | : | :
QY

```

NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California

COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 776 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-2

Query Match 1.88; Score 132; DB 1; Length 776;
Best Local Similarity 17.28; Pred. No. 0.0098;

Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 EENKLVGNLSMRIRTPDGGFFTRSAVOKFYODPKYGLIYNHD---NEYLYQDDWY--I 787
DB 72 EMYKAIGG---KIYVGDITKHSLEALSEDDKKIKDIYGDALLHEHYVAKGEIEYV 128
QY 788 LSSKVENSPEDYIFVYVYKGRNDAMDYGGSVLYTRSAVLPESIPELOTAAQVGRDFNT 847
DB 129 L---VIOSEEDYVENTERKALN-----VYVEIGKILSRDLISKINQYQKFLDVANT 176
QY 848 FIKTDNCG-----PEPPIYERLEKKEVEEERITIR----- 879
DB 177 IKNASDSDGODLFTNOLKHPIDFVSFELEQNSNEVEYFAFAVYIEPQHRDVLQLY 236
QY 879 -----EVEEIEEVEKVDKREVTLSKLFEGEKE----- 908
DB 237 APEAFNFMDFNDEIHLSELELKQDMLSRYEKMEIKQHYOHMSDSLSEGRGLLKL 296
QY 908 ---LQREBNFLKELKEEDVDLGLKMEATEV---EKLEGRALPI-----RKL 950
DB 297 QIPLPKKDDIHLISOEKEELKRIQIDSSDFLSTEKEFEKLQIDIRDSLSEEEKEL 356
QY 951 MAVAHCFSPCHRIAFESDDGIGRLGTRKRINGTEFLKLPRIQASD---LRTTG 1006
DB 357 L-----NRQYDSSNP-----LSEK--EKEFKKTKLDDIQPYDINORLQDTG 396
QY 1007 GRSSRPL-----SAFRSGFSKGIPIVLPSPKNEKLKELTAPLLKLTVG 1049
DB 397 GLIDSPSINLDVRYKQKRDIONIDALLHOSIGSTLVNKIYVENNMNINLTATL----- 451
QY 1050 VLACALIVPSAVALKATCACLKAGRIELAKCIANPACANVACIOTCNRPDETEC 1109
DB 451 -----GADLVDSJ----- 459
QY 1110 QIKCGDLFENSVDV---FNECAVSRKKCVPRK---SDLGEPAPDPSSVLYQNFNSDENG 1164
DB 459 -----DNTKINRGIFNFKNFYSSISNMIVDINERPLDERKLRQIKRQISPDTR 510
QY 1165 KWTYSGLNFPTDAFDQJLHEFTTEGDNKLY--GNISWRIKTLDGSGFFTRSAVOKFVQDP 1222
DB 511 AGYLENG-----KLIIQNRNIGLEIKIKVO----- 534

QY 1223 NOGVLYNHNDNEYLHYODDWYLLSKIKNKPEDYIFVYVYKGRNDAMDYGGAVYTR--S 1280
DB 534 ---LIKQSEKEYIRI--PAKVPRKSIDTKIOE-----AQLNINQMNKALLPRTKLTIT 584
QY 1281 SVLPNSIIELEKAKKSGIDGSTFIRTDNCGPEPALVERIEKTEVEGE--RIYKEY-- 1338
DB 585 FVHNHFRNYSNIVESAYLLINENKNNIQSD-----LIKRVNYLVYDNGRNFVFTDITL 636
QY 1338 -----EEIEEVEKEVEKVGRTMTLFLORLAEGFNLKODEENFVRE----- 1380
DB 637 PNIAEGYTHQDEIYQVASKGLYVPESSILLHGSKGV--ELRNDSEGFHFEPGHAVDY 695
QY 1380 ---LSKEEME-----FLDEIKMEASEV 1398
DB 696 AGYLLDKNOSDLVYNSKKRIFDKKEGSNL 725

RESULT 5

PCT-US94-01624-2
Sequence 2, Application PC/TUS9401624

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.

APPLICANT: Klimpel, Kurt R.

APPLICANT: Atora, Naveen

APPLICANT: Singh, Yogendra

APPLICANT: Nichols, Peter J.

TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:
ADDRESS: TOWNSEND and TOWNSEND, KHOURIE and CREW

STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94105

COUNTRY: USA

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624

FILING DATE: June 25, 1993

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-115

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 776 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

PCT-US94-01624-2

Query Match 1.88; Score 132; DB 3; Length 776;
Best Local Similarity 17.28; Pred. No. 0.0098;

Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 EENKLVGNLSMRIRTPDGGFFTRSAVOKFYODPKYGLIYNHD---NEYLYQDDWY--I 787
DB 72 EMYKAIGG---KIYVGDITKHSLEALSEDDKKIKDIYGDALLHEHYVAKGEIEYV 128
QY 788 LSSKVENSPEDYIFVYVYKGRNDAMDYGGSVLYTRSAVLPESIPELOTAAQVGRDFNT 847

```

Db 129 L---VIOSEDYVENTEKALN-----VYEEIGKILSRDILSKINQYOKFLDVINT 176
QY 848 EIKDNTGCG-----PEPLVERLEKKEVEGERTIIRK----- 879
Db 177 IKNASDSDGODLLFTNOLKEHPIDSEVEFLQONSNEVQVFAKAPAYITEPOHROVLQLY 236
QY 879 -----EVEEIEEVEKVRKREVTLLFSKLEFEGFKE----- 908
Db 237 APEAFNYMDKFNQOELINLSLEELKQRMLSREKWEKIKOHYQHSMSDLSSEGRGLKLKL 296
QY 908 ---LQDENFLELREKEMDVLDGLKMEATEV---EKLFGRALPI-----RKL 950
Db 297 QIPIEPKDDIHSLSQEEKELKRIQSDPLSTEKEFLKQIDIRDSLSEEEKL 356
QY 951 MAYATGCTSPCHDRIRFFSSDDGIGRLGTRKRINGTFLKLILPFIQSGAD---LRTG 1006
Db 357 L-----NRIOVSSNP-----LSEK---EKFLKLKLDIOYDINOQLQDIDG 396
QY 1007 GRSSRPL-----SAFRSGFSKGIFDIYLPBSKNELKELTAPLLKLTVG 1049
Db 397 GLIDPSINDVARKQYKRIQIDNIDALLHOSIGSTLYNKIYLVENMINMLTATL----- 451
QY 1050 VLACAFILPSADVAALKTACCLLGGRIELAKCIANPACANVACIOTCNRPDETEC 1109
Db 451 -----GADLVDSI----- 459
QY 1110 QIKGDLFENSVDY---FNECAVSRKKCVPRK---SDLGEPAPDPSVLQVNFIDFNG 1164
Db 459 -----DWTIKRIGFIENFKKFKYSISSNMYIDINERPLDNERLKLWRQLSPDTR 510
QY 1165 KWTITSGNLPTFADFQQLHEFTEGDNKLV--GNISWRIKLIDSGFFTRSAVQKVVQDP 1222
Db 511 AGYLENG-----KILQRIKIGLEIKDYQ----- 534
QY 1223 NQGVLYNHNEVLIHYODMYIILSSKIENKPEYIIVYVYGRDAMDYGAVVYTR--S 1280
Db 534 ---IKQSEKEYIRI--DAKVPRSKIDTKIOE---AQINIOENKAKALGPKRYKLLT 584
QY 1281 SVLPNSTIPELEKAASIGDFSTFRTDNTGCEPALVERIEKTEVEEGE--RIIVEV-- 1338
Db 585 FNVNHRVNASIVESAIIILNEMKNNIOSD-----LIKVTNYILVDGNGRVTFDIDL 636
QY 1338 -----EELIEEVEKEVEKVGRTENTLFLQRLAEGFENELKODENREVE----- 1380
Db 637 PNIAEQYTHODEIYEQVHSGLVPEPSRILLHGPRSGV--ELRNDSBGFHEFHGAVDX 695
QY 1380 ---LSKEEM-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNQSDLVYNSKRFIDFREEGSNL 725

```

```

RESULT 6
US-08-404-531B-6
; Sequence 6, Application US/08404531B
; Patent No. 5863724
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 5863724
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz MacLewicz &
; ADDRESSEE: No. 5863724fls
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,531B
; FILING DATE: 15-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1581 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-404-531B-6

```

```

Query Match 1.7%; Score 124; DB 2; Length 1581;
Best Local Similarity 18.8%; Pred. No. 0.15;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

```

```

QY 868 KVEEGERTIIEVEIEEVEKVRKREVTLLFSKLEFSGF--ELQDENFLRELSEKEMD 925
Db 480 KLSQNRFTTL-----EYSENERLKOT---NEMLRGILKILKIYAMENIFCSRYEKRRK 528
QY 926 VLDGLKMA--TEYKELFGRALPIRKIMAAVATHOFTSPCHDRIRFFSSD---GIGRL 978
Db 529 EMTSLRAFAVYTSISIEFNIAIPAAVLTFTV-----GHVSFFESPSVSVAASL 580
QY 979 GTRRRKINGTFL-----KILPIQ-----SADLRTTGSSRPLSAFSGSKG 1023
Db 581 SLFHLVLPFLPLSSVASTYKALVSQKLEFLLSSAIREEQCAPREPAPQGA---G 636
QY 1024 IFDIYPL-----PSKNELKELTAPLLKLVLGVLAFLVPSADVAALKTACCLLKG 1076
Db 637 KYQAVPLKVNKKRPAREVRDLGLPIQ-----RLTPSTDG--DADNFCVQIIG 684
QY 1077 C-----RIELAKCIANPACANVACIOTCNRPDEFECQIKGDLF 1117
Db 685 FETWTPDGIPLNSITTRIPRGQLMTIVGOVCGKSSLLATLG---EMQVNSAVF 738
QY 1118 ENSVDENFNECAVSRKKCVPRKSDLGEEFAPDP-----SVLVQNFNI--SDF 1162
Db 739 WNSLPD-----SEGRRRPOQPRAGDSGRFCQEQRPQCGYASOKPWLNAVTEINTFSPF 793
QY 1163 NGKWY---ITSGNLPTFADFQQLHEFTEGDNKLVGNISWRIKT----- 1205
Db 794 NKQRYKVIEMASLOPDIDIL-----PHGDOTOLERGINSITGGORPPDQCRPEPST 845
QY 1205 -----LDSGF-----FTRSAVQKVVQDPNQGVLYNHNEVLIHYODMYIILSS 1247
Db 846 STPMIVFLDDPFSALDVHLSHLMQAGILELDRDKRVVLYTHKLOLYLPA--DWITAMK 904
QY 1248 KIENKPEYIIVYVYGRDAMDYGAVVYTRSSVLPNSIIPLEKAKASIGRDFSTFIR 1307
Db 905 DGTIQREGTLKQFQSRSEQLFEHW-----KTIAMRQOQELK-----ETVME 946
QY 1308 TQNTGCEPALVERIEKTEVEGERIIVEVEIEEVEKE 1347
Db 947 RK---APEPS---QGLPRAMSSRDGLLDEDEEEBAESE 981

```

```

RESULT 7
US-08-404-531B-28
; Sequence 28, Application US/08404531B
; Patent No. 5863724
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel

```

TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
 Patent No. 5863724
 NUMBER OF SEQUENCES: 49
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 ADDRESSEE: No. 5863724 is
 STREET: One Liberty Place 46th. Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/404,531B
 FILING DATE: 15-MAR-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Beardsell, Lori Y.
 REGISTRATION NUMBER: 34,293
 REFERENCE/DOCKET NUMBER: BYLR-0003
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-568-3100
 TELEFAX: 215-568-3439
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1498 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-404-531B-28

Query Match 1.7%; Score 124; DB 2; Length 1498;
 Best Local Similarity 18.8%; Pred. No. 0.14;
 Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

868 KYEEGERITKEVELEEEVEKVRKKEVLFKLEEGK--ELQDENEFLELSKEMD 925
 480 KLSQAOITL-----EYSNERKQT--NEMLRGILKLYAMENIFCSREKTRRK 528
 926 VLDGLKMEA--TEVEKLRALPIRKLMVAATHCTSPCHDIRFFSSDD-----GIGRL 978
 529 EMTSLRAFAVYISISIFMTAIPIAVILTFY-----GVSFFKSDSPSAFASL 580
 979 GITRRKINGTELL-----KILPPIQ-----SADLRTGGSSSPPLSAFRSGFSKG 1023
 581 SLFHILVPLFLSSVSVSTKALVSVOKLSEFLSSAEIREBOCAPREPAPQQA---G 636
 1024 IFDIYPL-----PSKNELKELTAPLLKLVLGACAFLLVPSADANDALKTACLLKG 1076
 637 KYQAAFLAVNKKRRAREVRDLGPLO-----RLTPSTDG--DAONFCVQIIIGG 684
 1077 C-----RIELAKCIANPACAAVNAQLOTONNRPDETGCQIKGDLF 1117
 685 FETWTPDIGIPLTSLNITIRIPRGOLTMIVGQCGKSSLLATLG-----EMQVNSGAVF 738
 1118 ENSVVDENECVSKKCKVPRKSDIGEPFAPD-----SVLVQNFNT--SDF 1162
 739 WNSLPD-----SEGRRPQOPRAGDSGRFGCEQRPQGYASOKPMLNATVAEENITFESPF 793
 1163 NGKWY-----ITSGLNPTDFADQCQLHEFHTEGDNKLVGNISMRITK----- 1205
 794 NKQKRVKVIEMCSIQPDIDL-----PHGDQIQIEBERININSTGGQRPDQCRPEPST 845
 1205 -----IDSGF-----FTRSAVOKFVQDPNQPGLVYNHNDNEYLAHODDWYILSS 1247
 846 STPMIVFJDDDFSAIDVHLSHLMQAGILELRLDDKRTVYLVTHKQLPLPA--DWITAMK 904
 1248 KIENKPEYIFVYVYGRNDAMDGYGAVVYTRSSVLNPSIILELEKAKASIGRPFISFIR 1307

905 DGIQREGTLKDFORSECOLEFHW-----KTLNRODQLEK-----ETVME 946
 1308 TDNTCGPEPALVERIEKTEVEGERIITYKVELEIEVEKE 1347
 947 RK---APEPS--GQLPRAMSSRDGLIIDEDEEEEAASE 981

RESULT 8
 US-08-460-309-2
 Sequence 2, Application US/08460309
 Patent No. 5837496
 GENERAL INFORMATION:
 APPLICANT: Engvall, Eva
 APPLICANT: Leivo, Ilmo
 TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
 TITLE OF INVENTION: Fragments and Uses Thereof
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,309
 FILING DATE: 435
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/125,077
 FILING DATE: 22-SEP-1993
 APPLICATION NUMBER: US PCT/US 94/10730
 FILING DATE: 21-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/472,319
 FILING DATE: 30-JAN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/919,951
 FILING DATE: 27-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LA 9721
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1130 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-460-309-2

Query Match 1.5%; Score 115.5; DB 2; Length 1130;
 Best Local Similarity 20.0%; Pred. No. 0.46;
 Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;
 294 NEFLIYODDWYILSSQIENKPDYIFVYVYGRNDAMDGYGGSVITYTRSPLE--SIIPN- 352
 6 NDVKNEDHNLNGLKTRIENA-----DARNGLDLATLNDTLGKLSAIPND 49
 352 ----LQKAKSVGRFNNRITITDNSCGPEPLVERLEKTADEGKLLIKEAVEIEVEYEK 407
 50 TAAKIQ-AVKKARQAND--IAKDVLAQITTELHONLDGLKRNYNKL--ADSVARTNAVVK 104

Thu Oct 14 07:57:25 1999

us-09-075-375-2.raii

Page 7

```

Y 408 EVEK---VRPEMLT-----FORLEGFELOODENEFRELSEKEKELINLOMEATE 455
D 105 DPKSNKIADADATVKVLEDEADRLIDKLPKELEDBLNKKNS-ELKELINARQANS 163
Y 459 VE---KLFGALPIRLRLMALAPHSNELANHE- -IKYVGS-----KLPGH 500
D 164 IKAVSSGGDCIRTYKPEIKKGSNNINIVNKTAAVDNLLFYSKAFIDFLAIEMRKK 223
Y 501 KRFSWGM-----EDIESIVATICSRRIRPRFRKSPRCCGDSRGQ 545
D 224 VSEFMVDGSGVGRVEYEDLTIDDSYWRIVASTGRNGTL-----SVR--ALDGPKAS 274
Y 546 LESHGHNLSPAHSINONVPGKNSGCKFPKDVALMWEMKGOFAKTAIYAIFLVSASKA 605
D 275 IVPETHHSTP-----PGYITLDV-----DANMMLFPGGLTGL 308
Y 606 DAVDALKTCTCLLKECKLELAKCI-----SNPACAAVACLOTCNNRPETECQIKCGD 659
D 309 KRADAEVIT-----FTGCGETTFDNKP-----IQLMFREREKGCK----- 347
Y 660 LFENSVDDEFNECAVSRKKCVPRKSD-----VGDFPV--PPSVLYQKF-- 702
D 347 -----GCTVSS-----POVEDEBGTIQPDGEGYALVSRPIRKYPISTIVMKFR 390
Y 702 -----DMKDF-----SGKW-----FT 712
D 391 FSSSALLMYATRLRDLRPFMSVELTDGHIKYSYDLSGMSAVSNOHNDDKMSFTLSRI 450
Y 713 TRGALNPFEDAFDCOLHFEHFEENKLV-----GNLSWRIRTPD-----GGFTF-RSAVQKF-- 762
D 451 OKQANISIVDIDT-----NOENELATSSGNGNGLDLKADKTIYFGGLPLRLNLSMKAR 505
Y 762 -VQPKPYGILYNHNDNEYLLOYDWMYLLSSKVENSPEDYIFVYVYKGRNDAMDYGVGSVLY 820
D 506 EVNLKRTYSGCL---KQIEISRTPYNTLSS-----PDYGV-----TKGCSLENVY 547
Y 821 TRSAVLPESSIPELOTAQKYGRDFNFTIKDTNCG-----PPPLVE----- 864
D 548 TVS--FPRKGFVELSPVIDGTINISFSTKNSGILLGSGGTPAPPRKRRTQGOAY 605
Y 864 -----RLKKVEEGHRTIIRKEVEEIEEVEKVRD-KEYTLFSKLFESEFKELORDE-E 913
D 606 YVILLNKRLEVLHSTGARTRKRTIV--IRPPNLFPHGRHSVAVENTRGLITVQVDENR 663
Y 914 NLRLELSKEMDVLGLKMEATEVEKLF-GRALP 946
D 664 RYMONLIVE-----QPIEKKLFFVGAPP 687

RESULT 9
US-08-460-309-4
; Sequence 4, Application US/08460309
; Patent No. 5837496
; GENERAL INFORMATION:
; APPLICANT: Enyvall, Eva
; APPLICANT: leiyo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,309

```

```

1      FILING DATE: 435
2      CLASSIFICATION: 435
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER: US 08/125,077
5      FILING DATE: 3-SEP-1993
6      APPLICATION NUMBER: US PCT/US 94/10730
7      FILING DATE: 21-SEP-1994
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: US 07/472,319
10     FILING DATE: 30-JAN-1990
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: US 07/919,951
13     FILING DATE: 27-JUL-1992
14     ATTORNEY/AGENT INFORMATION:
15     NAME: Campbell, Cathryn A.
16     REGISTRATION NUMBER: 31,815
17     REFERENCE/DOCKET NUMBER: P-LA 9721
18     TELECOMMUNICATION INFORMATION:
19     TELEPHONE: (619) 535-9001
20     TELEFAX: (619) 535-8949
21     INFORMATION FOR SEQ ID NO: 4:
22     SEQUENCE CHARACTERISTICS:
23     LENGTH: 311 amino acids
24     TYPE: amino acid
25     TOPOLOGY: linear
26     US-08-460-309-4
27
28     Query Match 1.5%; Score 115.5; DB 2; Length 311;
29     Best Local Similarity 20.0%; Pred No.2,4;
30     Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps
31
32     QY 294 NEFLHYDDWYLLSSQIENKPDYIFVYRGRNDAMDGYGGSVIYTRSPLE-SLPN- 352
33         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34     DB 1987 NDVKEPNDHNLKTRLENA-----DARGDLRLTNDLTGKSAIPND 2030
35
36     QY 352 ----LQAAASVGRDFENFTTTNDSGCPPLVERLEKTAEGEKULLKEAVEIEBEK 407
37         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
38     DB 2031 TAAKIQ-VNVCVARQAND-TAKDVLAQITELHQNLDGLKKNYNL-ADSAKATNAVYK 2085
39
40     QY 408 EVER---VRDTEML-----FORLEGEKEQODENEFVRLSKEEKELINELQEMATE 458
41         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
42     DB 2086 DPKSNKIIAADAATVKNLBEADRLDLKLPKELEDNLKNIS-EIKELINQAOKQNS 2144
43
44     QY 459 VE---KLFGRLPLRKIRMAALPHSNFLANHET-----IKIYVGS-----KLPGH 500
45         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
46     DB 2145 IKVSVSSGDCIRITYKPEIKKGSYNNIVNVTAAVDNLFLYLSAKFIDELAIEMRKG 2204
47
48     QY 501 KRFSGW-----EDYGSIVYAKICSSRIPIRYFKSPRICGIDSRGLQ 545
49         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50     DB 2205 VSFIMDVSGSGREYREDLLIDSYWRIYASTGTNGTI---SVR--ALDGRKAS 2255
51
52     QY 546 LFSHGKHLSPNHSINONVPRKNSGCKFKPDVALMYERKMGQFAKTAIVAFILISVASKA 605
53         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
54     DB 2256 IVPSTHSTSP-----PGYTLIDV-----DANAMLEFGGLTGKL 2289
55
56     QY 606 DAVALKTCITCLKECKLDELAKCI---SNACACANVACLOTONNRDEFEQIKKGD 659
57         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58     DB 2290 KKADAVRIT-----FTGCGGETIYDNKP-----IGLMNFRKEBEDCK- 2328
59
60     QY 660 LFENSVDDEFETCAVSRKKCVPRKSD-----VGDPEV---PDPSVLVQKF- 702
61         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62     DB 2328 -----GVTS-----PQVEBSEGTIQFDGEGVALVSRPIRWPNISTVMEKRTI 2371
63
64     QY 702 -----DMKPF-----SGKM-----FI 712
65         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66     DB 2372 FSSSALLMYLATYRLCLRFPMSEVELTDGHIKVSYDLSGMAVSVNONHNDGKRKSFLLSRI 2431
67
68     QY 713 TRGLNPTFDATVYCOLHEFHFEENKLV-----GNLSWRIRPPD-----GFEFT-RSAVQKF- 762
69         : : : : : : : : : : : : : : : : : : : : : : : : : : : :
70     DB 2432 OKQANISITVIDT-----NOEENIATSSSGNNCGDLKKADKDIYTGGLTILANLSMKARP 2486
71
72     QY 762 -VQDKPYGIIYHNDNEYLLIYQDDWYLLSSKYKENSPEDIYFVYRGRNDAMDGYGGSVLY 820

```

Db 2487 EVNLKRYSGCL-----KDIEISRTPYNIILSS-----PDYVG-----TGCSLENNY 2528
 QY 821 TRSAVLPESTIIPLOQTAACKVGRDENTFIKTONTG-----PEPLVE----- 864
 Db 2529 TVS--FPRGPEVLSVPIDVGTETINLSTFKNESGIIILGSGGTAPPRRRRQTGOAY 2586
 QY 864 -----RLEKKVEGERIILKEVEEIEEVEKARD-KEVILFSKLFEGFELQDE-E 913
 Db 2587 YVILLNRGRLEVHLSTGARTMKIV--IRPEPNLFHDGRHSVHERTRGIFTYVDENR 2644
 QY 914 NPLRELKSEMDVLDGLKMEATEVEKLF-GRALP 946
 Db 2645 RYMONLIVE-----QPIEVKKLFYVGAP 2668

RESULT 10
 ; US-08-125-077-2
 ; Sequence 2, Application US/08125077
 ; Patent No. 5872231
 ; Patent No. 5872231 5840863
 ; GENERAL INFORMATION:
 ; APPLICANT: Engvall, Eva
 ; APPLICANT: Leivo, Ilmo
 ; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
 ; TITLE OF INVENTION: Fragments and Uses Thereof
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell and Flores
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/125,077
 ; FILING DATE: 22-SEP-1993
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US 94/10730
 ; FILING DATE: 21-SEP-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/472,319
 ; FILING DATE: 30-JAN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/919,951
 ; FILING DATE: 27-JUL-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Campbell, Cathryn A.
 ; REGISTRATION NUMBER: 31,815
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9001
 ; TELEFAX: (619) 535-8949
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1130 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-125-077-2

Query Match 1.5%; Score 115.5; DB 2; Length 1130;
 Best Local Similarity 20.0%; Pred. No. 0.46;
 Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;

QY 294 NEFLHYQDDWILSSQIENKPDYILFYVYRGRNDAMGYSVITNSPTLPE-STIPN- 352

Db 6 NDVAKENEDHLNGKTRILENA-----DARNGDLLRTINDTLGKSLAIPND 49
 QY 352 -----LQKAKSVGRDFENFTITDNSGPEPLVERLEKTAEBEGEKLILNEAVEIEEVEK 407
 Db 50 TAAALQ-AVADKARQAND--TAKDVLAQITELHQNIDGLKKNYKIL--ADSVAKTAAVAK 104
 QY 408 EVER--VRDTEMTL-----FQRLLEGKEILOQDEENFVRELSEKEEILNELQWATE 458
 Db 105 DPSKNKIADADATVKNLEQEADRLIDKLPKELEBDNLKNIS-EIKELINQARQANS 163
 QY 459 VE--KLFGRALPIRKRLMALPHSNFLANHEI-----IKYYGS-----KLPGH 500
 Db 164 IKVSVSSGGJCIRTYAPEIKKGSYNNIVNWKTAVDNLFLYLSAKFIIDPALEMRKK 223
 QY 501 KRFSGW-----EDYFGSIYVAKICSSRIPIRYFRSPRICGLDSRGQ 545
 Db 224 VSLFMDVSGVGVEYEDLDLIDSYWRYVASTGTNGTI-----SVR--ALDGPAS 274
 QY 546 LFSHGKHLSPASHINONVPKNGSGCKPRDVALMYERKGFAKTAIVAFILSVASKA 605
 Db 275 IVPSTHSTSP-----PGYTLIDY-----DANMLFVGLTGLK 308
 QY 606 DAVDAKTCCTCLKEGRLAKCI-----SNPACANACIOTCNRRPDETECOIKGD 659
 Db 309 KKADAVAVIT-----FTGCMGETTYDNKP-----IGLMFRKEEDCK----- 347
 QY 660 LFENSVYDEFNECAVSRKKCVPRKSD-----VGDFPV--PDPSVLVQKF-- 702
 Db 347 -----GCTVS-----PQVEDSEGTIQFDGVALVSRPIRMYNISTVWEKFT 390
 QY 702 -----DMQDF-----SGKW-----FI 712
 Db 391 FSSSALIMLYATRLDLRDFMSVELTDGIRKYSYDLGSGMAVSVNONHNDGKWSFTLSRI 450
 QY 713 TRGLNPTFDALDCQLHEFTEENKLV-----GNLSWRIRTPD-----GGFTT-NSAOKF-- 762
 Db 451 QKANSISYDIDT-----NOENIATSSGNNFGDLKADDDIYGGCLPTLNLMSKARP 505
 QY 762 -VODPYRPGILYHNDVEYLLYODWYIILSKVNSPEDEYIFYYKGRNDAMGYSVLY 820
 Db 506 EVNLKRYSGCL-----KDIEISRTPYNIILSS-----PDYVG-----TGCSLENNY 547
 QY 821 TRSAVLPESTIIPLOQTAACKVGRDENTFIKTONTG-----PEPLVE----- 864
 Db 548 TVS--FPRGPEVLSVPIDVGTETINLSTFKNESGIIILGSGGTAPPRRRRQTGOAY 605
 QY 864 -----RLEKKVEGERIILKEVEEIEEVEKARD-KEVILFSKLFEGFELQDE-E 913
 Db 606 YVILLNRGRLEVHLSTGARTMKIV--IRPEPNLFHDGRHSVHERTRGIFTYVDENR 663
 QY 914 NPLRELKSEMDVLDGLKMEATEVEKLF-GRALP 946
 Db 664 RYMONLIVE-----QPIEVKKLFYVGAP 687

RESULT 11
 ; US-08-125-077-4
 ; Sequence 4, Application US/08125077
 ; Patent No. 5872231
 ; Patent No. 5872231 5840863
 ; GENERAL INFORMATION:
 ; APPLICANT: Engvall, Eva
 ; APPLICANT: Leivo, Ilmo
 ; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
 ; TITLE OF INVENTION: Fragments and Uses Thereof
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell and Flores
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA

```

:
: ZIP: 92122
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/125,077
:
: FILING DATE: 27-SEP-1993
:
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US PCT/US 94/10730
:
: FILING DATE: 21-SEP-1994
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 07/472,319
:
: FILING DATE: 30-JAN-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 07/919,951
:
: FILING DATE: 27-JUL-1992
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Campbell, Cathryn A.
:
: REGISTRATION NUMBER: 31,815
:
: REFERENCE/DOCKET NUMBER: P-LA 9721
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (619) 535-9001
:
: TELEFAX: (619) 535-8949
:
: INFORMATION FOR SEQ ID NO: 4:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 3111 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: US-08-125-077-4

```

Query Match 1.5% Score 115.5; DB 2; Length 3111;

Best Local Similarity 20.0% Pred. No. 2.4;

Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;

```

QY 294 NEFLHYODDWYLLSSQIINPKPDYIFVYVYRGNDAMDYGGSVIYTRSPLE-SIIPN-352
DB 1987 NDVKNEDHLNGLKTRINA-----DARNDDLRLTNDITGKLSAIPND 2030
QY 352 ---LQAKASVGRDNFNFTTNSGCEPPLVERLEKTAEGEKLIRAEVEEVEVER 407
DB 2031 TAAKIQ-AVKDKARQAND--TAKDVLAQITELHQLDGLKKNYKL--ADSVANTNAVVK 2085
QY 408 EYER---VROTEMIL-----FQRLLEGFKELQODEENFVELSKKEEITINELQEMATE 458
DB 2086 DSKKKRIADADATVKNLEQADRLIDKLPIKLEDMIKKNIS-EIKELINQARKQANS 2144
QY 459 VE---KLFGRALPIRKLMLALAPHSNFLANHET-----IKYYVGS-----KLPQH 500
DB 2145 IKVSVSSGGDCIRTKPKPIKGSYNNIYVKNIAVADMLLYLSGAKTIDFLAEMRGK 2204
QY 501 KRFSGW-----EDYFGSIYVAKTSSRRIRIPRYFKSRPICCGGLDSRLQ 545
DB 2205 VFLMDVSGVGRVEYEDLTIDDSYWRIVASRTGRNGTI-----SVR---ALDGPKAS 2255
QY 546 LFSHGKHNISPAHSINQNPVKNKSGCKPRKVVALVWKEKMGOFATLVAFLILSVASKA 605
DB 2256 IVPSTHHSSTP-----PGYTILDV-----DANAMLFVGGTLGKL 2289
QY 606 DAVDALKTCTCLKECRLELAKCI-----SNPACAAVACLOTQNNRPDETEQIKGCD 659
DB 2290 KKADAVRVI-----FTGCMGEITYFDNKP-----IGLMNFKREKGDCK----- 2328
QY 660 LFNENVDFEFCNAVSRKKCVPRKSD-----VGDFPV---PDSVLYQKF-- 702
DB 2328 -----GCTVVS-----POVEDSEGTIOFDGEGYALVSRPIRWYPNISTVAFKFR 2371
QY 702 -----DKKDF-----SGKW-----FI 712
DB 2372 FSSSALLMTLATRLDLRDFMSVELTDGHIKVSYDLGSGMASVAVSNQNHNDGKRSFTLSRI 2431

```

```

QY 713 TRGNIPTFADPFCQQLHEFEETENKLV-----GNLSWRIRTPD-----GGEFT--RSAYQKF-- 762
DB 2432 QKQANISYIDIDT-----NQENIATSSGNNFGDLADKITYGGIPLIRNLMSMKRP 2466
QY 762 -VODPYPGILYNDNEYLLYDDWYIILSSKVENSPEDYIFVYVYGRNDAMDYGGSVLY 820
DB 2487 EVNLKRYSCCL-----KDIEISRTPYNIIS-----PDYGV-----TKGCSLENY 2528
QY 821 TRSAVLPEESILIELQTAQKAGRDENFTIKDNTG-----PEPPIVE----- 864
DB 2529 TVS--FPRPGYVELSPVPIDVGTETNLSTKNSGIIILGSGGTPAPRRRRROTGCAY 2586
QY 864 -----RLFKVEEGERTIKVEEIEEVEKVRD--KEYTLFSKLFEGFKELQORDE-E 913
DB 2587 YVILLNRGLFVHLSTGARTMKRV--IRPENLFDGGRHSVHERRTGRTITVQVDENR 2644
QY 914 NPLRELKSEMDVLDGLKKEATEVEKLF-GRALP 946
DB 2645 RYMONLIVE-----QPIEVKKLFVGGAP 2668

```

RESULT 12

US-08-568-459A-4

Sequence 4, Application US/08568459A

Patent No. 5849306

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

ADDRESS: Knobbels Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/568,459A

FILING DATE: 07-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelisen, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121,001CPI

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1435 amino acids

TYPE: amino acid

STRANDEDNESS: single

MOLECULE TYPE: protein

HYPOTHEICAL: NO

ORIGINAL SOURCE:

ORGANISM: Plasmodium falciparum

US-08-568-459A-4

Query Match 1.5% Score 115; DB 2; Length 1435;

Thu Oct 14 07:57:25 1999

us-09-075-375-2.rai

Page 10

Best Local Similarity 19.3%; Pred. No. 0.75;
Matches 171; Conservative 122; Mismatches 315; Indels 280; Gaps 46

QY	552	HNLSJPAHS	INOVUPGNSGCKEPPKQVLA	WMXWKGCFKTIIVAFITLSA	KADAVDAL	611				
Db	378	HILSNYE	-TQVKEEN	-----	AENLIT	-SENKRAKAVSL 412				
QY	612	KTCITLLKECRLEIAK	CI	SNPACAANVACLOT	CNNRDETECOQICGDLFENS	YVEFN- 671				
Db	413	-----	LNNCAEVS	KYCD	-CKHTITLVSV	INGNDNT---IK-EKREHIDDFSK 459				
QY	671	-----	EC	-----	AVSRKCYR	-----				
Db	460	FGCDKNSVD	NTKWECKNPI	ILSTIKDYCV	PROEOLC	IGNDIRYKNLMTREHLAT 519				
QY	690	PVPDDSVLVOK	EDMKDFSGKWEIT	RGLNPTPEDAF	DCOLHEFHTEENKVLNLS	-WRITRP 748				
Db	520	AIYESRI	LKRKYKNDKE	-----	VCKIINKEP	-----				
QY	749	DGGFTTSR	AVOKFVODPYR	PGILNHNH	EYLLIYODDWI	ILSKYENSPEYITVYIKGRN 808				
Db	561	-NDLSNR	VLGKINTNSKY	-----	-VHRNKNKDLF	EDWKVIYK----- 600				
QY	809	DAMDYGSG	SVLYTRS	ASVLPESII	PELQIAAKVGRD	ENTFIK-TDNTCGEP	PLVERL-- 866			
Db	600	DVMNVI	-----	SWFVKOTV	KEDDI-----	ENIQFFRWS	SEWMDVCOQRTKMIETIKY 650			
QY	866	---EKV	EG--	-----	EPTIKVE	VELEBEVEYKVRDEVTLSKLEFGE	LODEE 913			
Db	651	ECKEPP	CCDDNCKSK	NSYKFM	IKSKKEEYKNO	KQOYEOYKGNMY	SEFSIK--BE 708			
QY	914	NELRELS	---EMOY	DLTKMEATE	VEVELEFRALP	PIKILMAVANTHCT	SHCHRIFFSS 971			
Db	709	YVLKYS	SKCSNLN	PEDEKFEL	SDYK-----	NKTCMP	----- 744			
QY	972	DDGIGRL	IGTRKR	INGFTFLKIL	PIPSADRT	ITGRSSRPLSAFRSGSKG	IFDIPLP 103			
Db	744	---EY	CDVISTIR	NEQTS	CPAVE	-ENTEI--	AHRETP-----SISEG-----P 784			
QY	1032	SKNELKEI	TAPLLK	INGVLA	CALIYP	-----	SADAVDLKTCACI-LKG-CRIFLACIANM 108			
Db	785	KGNKEKE	DDDSLKI	-----	SVSPENS	RPTDCKD	ISNLKLGVDIMPR	AVIG 836		
QY	1088	PACAANV	ACLOT	CNN-----	RPDETCO	QICGDLFENS	VDFEFCASRRK 113			
Db	837	SSPMDN	INIVTEQ	GNISGVNSKPL	SDVDPR	PKLEEDQ	NSDEBEETVYVN-----ISKSP 891			
QY	1135	CYPRKSD	IGEP	PADPSV	LVONFIS	-----	DENCKWIT-TSGLNPT	DAFDCQHEFHTEG 119		
Db	892	SINNGD	SGSASV	SSSSSNTOL	SIDDRNG	DITFRTD	ANTEDV-----IRKENADK 947			
QY	1191	DNKLVGN	ISWRIK	TLDSOFF	RSR	AVOKFODPN	PGVLVHND	EYHJ-----XODDWI 124		
Db	948	DEDEK	GADEH	HSSES	---LSSE	EEMILD	-NEGNSL	NHEEYKHTS	NSDNVOOGSI 100	
QY	1245	LSSKIE	KNPE	DIYIFVY	YRGR	DWADWYG	AVYVYTR	SSVINSIIP	ELKAKRSIG	DFST 130
Db	1004	VNANV	VELK	DTL-----	ENPS	-----	SLDEGR	AHEEL	SEP	NSISSDDMS- 104
QY	1305	FIR	DN	TCG	PEP	ALV	ERIEKTVE	-----	EEERIV	VEVEI 1340
Db	1045	-----	NTPG	-----	LDNT	SEET	ETIS	INNEYK	VNERDE	ERTITVEYDI 1084

```

1 TITLE OF INVENTION: AND METHODS OF USE
2 NUMBER OF SEQUENCES: 4
3 CORRESPONDENCE ADDRESS:
4 ADDRESS: DANN, DOFFMAN, HERRELL AND SKILLMAN
5 STREET: 1601 MARKET STREET, SUITE 720
6 CITY: PHILADELPHIA
7 STATE: PA
8 COUNTRY: USA
9 ZIP: 19103-2307
10
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: PatentIn Release #1.0, Version #1.25
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/08/353,700
18 FILING DATE: 09-DEC-1994
19 CLASSIFICATION: 435
20 ATTORNEY/AGENT INFORMATION:
21 NAME: REED, JANET E.
22 REGISTRATION NUMBER: 36,252
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (215) 563-4100
25 TELEFAX: (215) 563-4044
26 INFORMATION FOR SEQ ID NO: 1:
27 SEQUENCE CHARACTERISTICS:
28 LENGTH: 3248 amino acids
29 TYPE: amino acid
30 STRANDEDNESS: single
31 TOPOLOGY: linear
32 MOLECULE TYPE: protein
33 HYPOTHEetical: NO
34 ANTI-SENSE: NO
35 ORIGINAL SOURCE:
36 ORGANISM: HUMAN
37
38 JS-08-353-700-1

```

Query Match	1.5%;	Score 112.5;	DB 1;	Length 3248;
Best Local Similarity	17.7%;	Pred. No. 4.6;		
Matches	219;	Conservative 178;	Mismatches 408;	Indels 431;
			Gaps	60

```

QY 381 ERKTAEB--GKKLLKEVLEEEVEEVEKVRD---TEMTLQORL-----EBCF 426
Db 883 QRIKLOEDTSÄONVVAETLSALENKEKELOLNDKYTEBOALIOELKSNHLLSDSK 942
QY 429 ELAODEENFVELSKKEKETLNDLOMEATEVEKLGFRALPIKTLMAALPHS--NFLANHE 487
Db 943 ELQLLSTET--LSEKKEKMSITSLNKRREIBELQENLTKEINASLNQEKEMNLIOKSE 998
QY 488 TIKYVSKLPGKRRFSWGNEDYFGSIVAAKICSSRRIPRIFRKSPRIICGLDSRGLQF 547
Db 999 SFANYT-----:-----DER-----1008
QY 548 SHGHNLSPASHINONVPGKNSGCRPPQVALMWEMKMGQFAKTAIVAFILSVASKADA 607
Db 1008 -----EKSISELSDYOKKEKILLQORCEFTGN-----AYEDLSQKYKA-A 10466
QY 608 VDLAKTCTCLFKRCRLAKLCISNPACAANVALOTCNNRPDETCQIKGDDLENSYVD 667
Db 1047 QEKNSKLECLLNE-----CTSLCENKRNKELE-QLK--EAFKHEQ 10844
QY 668 EFNNCASVRKCCPRKRSVDGDFPVDPVSLYOKFPMKDFSGKWFITRBLNPTLPDAFCQL 727
Db 1085 FLTKLFAEER-----NONLMLE-----LETVOQLRSEM 11144
QY 728 HEHTEENKLVGHLISWRIRTPDGGGFTFSRVAQFODEKYPGILYHNHNEILLYQDDWYI 787
Db 1115 TDNONNSKSEAGLQKEIMTLKE--EQKMKQKEVND-----LQENBQLM-----KV 11599
QY 788 LSSKVE--NSPEYIYVYKGRNDAMDYGGSVLYTRSVALPESTIPELQTPAOKRGVD- 845
Db 1160 MKRRKHEQNLIESPT-----RQ-----SVKREERENOCNKRPMQMLEVEISLDS 1205

```

```

OY 845 FNT-F1RTDNTCGPEPLVRLKXVEEGRITIKVEEILEEVEKVR-DKEVT-----LE 898
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1206 YNAOLVLE-----AMLNKRELKIOESE---KEKELJOHELQITRODLETSLQDMQ 1234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 899 SKLEGEKELORD-EENFL---RELKSEED-----VLQGLKMEATEVEKLFGRALPIR 948
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1255 SOEISGLKDEIDAEKXIYISGPEHLESTQDANHLCQSQTMTNKNLNELEKIC-ELIOAE 1313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 949 KLMVA-----THCFIS-----PCHDRIRFESSDDGIGRLGCTRKRINGTEFLLIK 993
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1314 KYEVLTELANDRSRCITATRKMAEEVGKLINEKYLINDSGL-----LHGE-LVED 1363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 994 LPPIQSADLRITG-----RSSRPISAFRSRSGSKGIDIVLPSKNEKLETAPLLKL 1047
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1364 IP-----GGEFEOPNEQHPVS-----LAPLDESNTYEHLT--LSQKE 1399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1048 VGV---LACAFLLVPS-----ADAVDLATCACLKGRICELAKCIAN----- 1088
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1400 VQMHFAELQEKFLISQEHKILHDQCHQMSKXKSELQITVYDSLKMNELVLSNLNFPQD 1459
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1088 -----PACAAVACLQTCNNRPETCQIKCGDL-----FENSVYDE 1124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1460 LYKEMQGLEBGLVPSLSSSCVPSDSSLSLSSGFYRALLBQTDMSLSLNLGAV--S 1517
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1125 FNECAVSRKKCVPRKSDLGEPAPDPDSVLYQVNPNISDFNG-----KWTYISGLNPEPDA 1178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1518 ANQCSVVEVFCSSIQITVDSLKAE--LVLSNLNFPQDGLYKEMQGLEBGLVPSLSS 1574
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1179 F-----DCQIHEFHTF--GDNKLVGNISMRK--TLDSGEFT----- 1212
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1575 SCVPDSSSLSSLGDSFYRALLBQTDMSLSLNLGAVSANQCSVVEVFCSSIQEENLTR 1634
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1212 -----RSANQKFPQDPQNPQVVLNHNNEVLNHOQWYILSKATE 1230
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1635 KETPSAPAKGVEELESICEVYRQSLERLEKEMESQIMKNKRIQELE-----QLTSS--E 1687
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1251 NKPEDIYFVYVYRGRNDAMDQYGAVVYTRRSVL-----PNSIIPLEKAAXSI-GRDPS 1303
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1668 RQELDCRLKQYLSNEMQWOKLISVLLNEMSKLAAKKQTEBLSLEFVARQLQGLDLS 1747
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1304 TFIPTDNTCGPEPALVERIK-----TVEGGERIIVKVEEI-----EEBEVEKEKVR 1351
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1352 GRF-----EMTLFQRLAEGENL-----KQ 1371
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 1806 TETAVKPTGCGSEGSEPTNYTPPPGHDKI--QGSSECISELSTFSGPNALVPMDFLGNQE 1863
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1864 DHHNLQRLVKETSNENRLLHVEDRDKRKESSLNE 1899
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
PCT-US95-16216-1
: Sequence 1, Application PC/TUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yen, Timothy J.
: APPLICANT: Ratner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Hettrich and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

1      OPERATING SYSTEM:  PC-DOS/MS-DOS
2      SOFTWARE:  Patent In Release #1.0, Version #1.30
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER:  PCT/US95/16216
5      FILING DATE:
6      CLASSIFICATION:
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER:  US 08/353,700
9      FILING DATE:  09-DEC-1995
10     ATTORNEY/AGENT INFORMATION:
11     NAME:  Reed, Janet E.
12     REGISTRATION NUMBER:  36,252
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE:  (215) 563-4100
15     TELEFAX:  (215) 563-4044
16     INFORMATION FOR SEQ ID NO: 1:
17     SEQUENCE CHARACTERISTICS:
18     LENGTH:  3248 amino acids
19     TYPE:  amino acid
20     STRANDEDNESS:  not relevant
21     TOPOLOGY:  not relevant
22     MOLECULE TYPE:  protein
23     HYPOTHETICAL:  NO
24     ANTI-SENSE:  NO
25
26     PCT-US95-16216-1

```

[illegible]

```

Db 1314 KYELVTELNDSRSECIATATKMAEYVGLNLEVKILNDSGL-----LHGE-LVVD 1363
QY 994 LPLQSDADLTG-----RSSRPLSAFRSGSGSKIFDIPLPSKNELKELTAPLLKL 1047
Db 1364 IP-----GGEGEOPNEOHVPS-----LAPIDESNSEYHET--LSDKE 1399
QY 1048 VGV-----LACAFILVPS-----ADAVALKTCACILKGGRIELAKIAN----- 1088
Db 1400 VQMHFALQKFLSLQSEHKILMDQHQMSKMSKSELQTYDSIKAEINVLSTLRNFGQD 1459
QY 1088 -----PACANVACLOTCNNRPDETECOIKGDU-----FENSVD 1124
Db 1460 LKEMOLGBEGLVPSLSSCVPDSSSLSDGSFRRALLEQTMSSLSNEGAV--S 1517
QY 1125 FNEGAVRRKRCVPRKSLGFEPPADPSVLYONFNISDFNG-----KWITSGLNTPDA 1178
Db 1518 ANOCSDVEVFCSSLOQTVDSLKAEN--LVLSTLRNFGQDLVKEMQLGEBGLVPSLS 1574
QY 1179 F-----DCQLHEFHT--GPNRLVGNISWRK-----TLDSGFT----- 1212
Db 1575 SCVPDSSSLSDGSFRRALLEQTMSSLSNEGAV--S 1634
QY 1212 -----RSVQKQVODPNQPGVLYNHDNELYLYODWYILSKIE 1250
Db 1635 KETPSAPAKGVLEELSECEVYRQSLKLEKEMESOGIMKKEIOELE-----QLSS--E 1687
QY 1251 NKPEDYIFVYRGRNDAMDYGAAYYTRSSVY-----PNSITPLEKAKAI--GRDS 1303
Db 1688 RQELDCIRKQYLSFENEGWQOKLTVLEMSKLAEEKQTEQLSLLEVARLOQLDLS 1747
QY 1304 TFIKTDTCGPEPALVERIK-----TVEEGERTIVKEVEI-----EEVEKEVEKY 1351
Db 1748 S--RSLGIDTBDALQGRNSCDSISKHETSETERTPKHGVHICCKDAQDQDINDIEKI 1805
QY 1352 GR-----EMTLFQRLABGFNEL-----KQ 1371
Db 1806 TETGAVKPTGECGSGEOSPDTNYPEPGDKT--OGSSSECLSELFSFSPNALVPMDFLNGOE 1863
QY 1372 DEENF--VRELSEKEEFLEDEIKMEASEVEKLEFGK 1404
Db 1864 DIHNLQLRVKEITSNENIRLLHVIEDRKRKVESLNE 1899

```

RESULT 15
US-08-603-753D-4
Sequence 4, Application US/08603753D
Patent No. 5891857

GENERAL INFORMATION:

APPLICANT: HOLT, JEFFREY T.

APPLICANT: JENSEN, ROY A.

APPLICANT: PAGE, DAVID L.

APPLICANT: KING, MARY-CLAIRE

APPLICANT: SZABO, CSTILA I.

APPLICANT: JETTON, THOMAS L.

APPLICANT: ROBINSON-BENION, CHERYL L.

TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2

TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON

TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARLES A. TAYLOR, JR.

STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER

CITY: DURHAM

STATE: NORTH CAROLINA

COUNTRY: USA

ZIP: 27707

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 800 KB storage

COMPUTER: IBM PC/XT/AT compatible

OPERATING SYSTEM: Windows 3.1

SOFTWARE: WORD PERFECT 6.1 and ASCII

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603,753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-08-603-753D-4

Query Match 1.5%; Score 111.5; DB 2; Length 3418;
Best Local Similarity 17.8%; Pred. No. 6.1; 576; Indels 483; Gaps 72;
Matches 275; Conservative 212; Mismatches

QY 3 LSLTFVCKEELNLYARSPCNERFRHSGOPT---NIIMKIRSNNGYENFRLETSY 59
Db 540 LEITVCSQKEDSL-----CPULIDNGSWPATTONNSVALK---NAGLISLTKRTNK 589
QY 60 KTSFSDSSSHCK-----DKSQICSIDTSFEEIQREFDKRGMTLILEKQWFOFLAIV 112
Db 590 FIYAIHETEFYKGRKIPKDOSEILNCSAF--EANAFE---APLIFANADSGILHSSVY 644
QY 113 LVCTFVIVPRVDAVDAKTCACILKEKRILAKCIANPSCAAVACLOTCNNRPDETEQ 172
Db 645 RSCS-----QNDSEPTLSLTSFGTILKRCSSRETCSNNTVIYSQDLDYK--EAKCN 694
QY 173 IKCGDLFENSVD-----QFNECAVSRKRCVPRKSDVGEFPV-----PDRNAVQNFNMK 222
Db 695 KEKIQLTTPRPAVDSLQSGQCEKNDPKS--KRYSDIKKEVYLAACHPVQHSKVF--YSDT 751
QY 223 DFGSKWTI-----TSGINPTFDADFQD-----HEFHMEKDKLVGNLTWRITLDG 268

```

```

Db 752 DFOSQSLYDHENASFLIIPISKDVLSNLVMSRCKESYKMSDKLGN----- 802
QY 269 GEFIRSAVOTFEVODPD.PGALYNHNDNEFLHYODDWYLLSQIEN---KDDYIFVYRG 324
Db 802 -----NESVELTKNIPMEKN-----DVCALNENYKNVLLPEKIMRPAASPS 846
QY 325 RNDAMDYGGSVYIYRSPLEPIESIIPLQAKANSVGR-----DENNETI----- 368
Db 847 R-----KVQENQNTNL--RVIOKNQEBETISISKITVNPDESELLFSENENNFFQOVA 895
QY 368 -----TIDNSCGPEPLV-ERLEKTAEEGK-----LLIENAVEIEEVE 406
Db 896 NERNNMLAGNTKELHEDLTCVNEPIFKNSTMYLYGDTGDKQATQVSIKRDLYVLAEBN 955
QY 407 KEVEKVDTEMTLEFORL-----LEGFKELQDEEN-----FVRELK 443
Db 956 KNSYK-OHIKWTLGQDLKSDISLNDIKIPEKNNDYMKKAGLIGPISNHSFGSFRJASN 1014
QY 444 EKEEI-----INELQEMATEVEKLEGRALPIRKLMLALA-PHSNPLANHETIKYVGSKL 498
Db 1015 KEIKLSEHNKRSKMFEDIEQYPTSLACVEIYNTLALDNQKKLSKQSI-----NTVS 1069
QY 499 GHKRFSGWEDYFGSIYVAKICSSRRIPRYFRKSPRICGLDSKGLDLS-----HGKHN 554
Db 1070 AHLQ-----SSVYVSDCKNSHITPO-----MLFSKODFNSNHL 1103
QY 555 SP---AHSHINONVPKNSGCKPKDVALMWKMGOFKTAIYAFILSVAS---KADAV 608
Db 1104 TPSQKAEITELSTLIESGSP-----EFTQFRKS---YLOKSTFEVPEKNOM 1149
QY 609 DALKTCTCLKECRLEAKCISNPACANVACIOTCNNRPEDECOITKCGDLFENS--- 665
Db 1150 TILKTS---EBCRDADLHVIMN---APISQVDSKQEBGVEIKRRFAGLLKNDCKNS 1203
QY 665 ---VDE-----FNECAVSRKKCVPRKSDYGDPPVDPVLYOKFKMD 705
Db 1204 ASGYLTJEDENEYGFEGFYSAHGTRKLVNSTEALQKAVKLEFSDIENIS-BETSAEVAHPISS 1262
QY 706 FSGKWFITRGILNPFDAFDCOLHEFHT---EENKLVGNLSWRIRTPDGGFFTRSAVOKFV 762
Db 1263 -----SKCHDSVYSMEKIEINHNDKTYSEKNKCOLILONNIMTGTGTF---VEIT 1310
QY 763 QDPKYPGILYNHNDNEFLYODDWYLLSKVENSPEDEYIFVYKGRNDAMDYGGSVLYTR 822
Db 1311 EN--YKRNTEHEDNKYTAASRNSHNLERFDSSKNDTVCIHKDETD-----LIFTD 1360
QY 823 SAVLPESIIPELOTAOKVGRDFTFKTD-----NTCGPEPLVER 864
Db 1361 Q-----HNICLKLISQOFMKEG---NTQIKREDLSLTFLEVAKAOEACHGNTSNKEQLATK 1413
QY 865 LEKVEEGERTIIEKEVEIEEVEKVRDKEVT---LFSKLEGEFKELQDEENFL----- 917
Db 1414 TEQN-----IKDEFISDTFFOTASGKNISVAKELFNKIVNFDDOKPELHNFLANSE 1465
QY 917 --RELSEKMDVLDGLKMEATEV--EKLFGRALPIRKLMAVATHCFTSPCHDIRPFSSD 972
Db 1466 LHSDIRKKMDI---LSYEETDIYKHKLIKESVPV----- 1498
QY 973 DGIGRLGIT-----KRRINGTFL-----KILPPIOSAD---LRTTGRSSRPL 1013
Db 1498 -GTGNOLVTFOGOPERDEKIEPTLLGFHTASGKVKYAKESLDKVNLLFDEKEGOTSEI 1556
QY 1014 SAFRGSFGSGJFDIYPLPSKNEKLELTAFLKLVLGVLACAFLLVPSADAVDALKTACL 1073
Db 1557 TSFSHQWAK-----TLKYREACKDLE-----LACETIETITAAPCKEOMON--SL 1598
QY 1074 LKGCRIELAKCIANPACANVACIOTCN-----NRPDET-ECQITC----- 1114
Db 1599 NNDKRLVSIETVVPKLLSDNLCRQTEHLKTSIFLKVKVHENVEKETAKSPAFCYTNO 1658
QY 1114 --GDLFENSVDDEFNECAVSRKKCVPRKSDGLGEPAPDPVLYONFNISDFNGKWTITSG 1171
Db 1659 SPYVSIENSALAFYTSK--SRKTSVTSQTSLL-----EAKKWLREG 1696

```

```

QY 1172 INPTEADPCQLHEFHTEGDNKLVGNISWRITKTLDSGFFTRSAVOKFEVODPNQGVLYNH 1231
Db 1697 I-----FDGQPE-----RINTAD-----YGNVLYENNSSTIAEN 1727
QY 1232 DNEVLYHODDWYLLSKKLEKNPEDEYIFVYRGRNDAMDYGGAAYVYRSSVLPNSIIPEL 1291
Db 1728 DKNHLSKQDITYLSNSSMSNS-----YSYHSDDEVYNDSG-----YLSKNKLDGIEPVL 1776
QY 1292 EKAASIGRDPSTFIRTDNTCGPEPALVERIEKTYEEGERIYKEV 1337
Db 1777 KVEDOKNTSFESKVISNVKANAPQTV-----NEDICVEEL 1813

```

Search completed: October 14, 1999, 00:46:49
Job time: 740 sec

Thu Oct 14 07:57:27 1999

us-09-075-375-3.rml

Page 1

GenCore version 4.5
Copyright (c) 1993 - 1998 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 14, 1999, 00:42:50 : Search time 83.2 seconds
(without alignments)
1987.079 Million cell updates/sec

Title: US-09-075-375-3
Perfect score: 1589
Sequence: 1 tttttcagtggttgcagt.....tctttcagaagcaaaaa 1589

Scoring table: IDENTITY_NUC

Searched: 192659 seqs, 52021692 residues

Database : Issued Patents, NA:*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/5C.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/5D.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCUS9.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	75.8	4.8	7218	2	US-08-232-463-14 Sequence 14, Appl
2	40.2	2.5	3092	2	US-08-426-627-3 Sequence 3, Appl
3	40.2	2.5	3253	2	US-08-426-627-5 Sequence 5, Appl
4	38.8	2.4	3337	1	US-08-072-610-1 Sequence 1, Appl
5	38.8	2.4	3337	1	US-08-719-822B-1 Sequence 1, Appl
6	38.2	2.4	2277	2	US-08-676-967-2 Sequence 2, Appl
7	38.2	2.4	2277	2	US-08-676-974-2 Sequence 2, Appl
8	38.2	2.4	2277	2	US-09-098-487-2 Sequence 2, Appl
9	36.6	2.3	1052	2	US-08-466-603-1 Sequence 1, Appl
10	36.6	2.3	1052	2	US-08-314-503A-1 Sequence 1, Appl
11	36.6	2.3	1052	2	US-08-468-066-1 Sequence 1, Appl
12	36.6	2.3	1052	4	US-08-466-717-1 Sequence 1, Appl
13	36.6	2.3	1052	5	PCR-US95-12414-1 Sequence 1, Appl
14	36.4	2.3	5893	3	US-08-592-126-54 Sequence 54, Appl
15	36.2	2.3	24701	4	US-08-853-659A-2 Sequence 3, Appl
16	36.2	2.3	24701	4	US-08-853-659A-3 Sequence 3, Appl
17	36.2	2.3	15512	4	US-08-853-659A-5 Sequence 5, Appl
18	36.2	2.3	15512	4	US-08-853-659A-8 Sequence 8, Appl
19	36.2	2.3	24701	4	US-08-853-659A-60 Sequence 60, Appl
20	36.2	2.3	24701	4	US-08-853-659A-61 Sequence 61, Appl
21	36.2	2.3	15512	4	US-08-853-659A-63 Sequence 63, Appl
22	36.2	2.3	15512	4	US-08-853-659A-66 Sequence 66, Appl
23	35.6	2.2	2223	3	US-08-257-073-4 Sequence 4, Appl
24	35.4	2.2	797	3	US-08-752-132-1 Sequence 1, Appl
25	35.2	2.2	188	1	US-08-115-497-21 Sequence 21, Appl
26	35.2	2.2	188	3	US-08-466-670-21 Sequence 21, Appl
27	35.2	2.2	188	4	US-08-291-011-1 Sequence 1, Appl
28	35.2	2.2	405	4	US-08-299-074A-1 Sequence 1, Appl
29	34.8	2.2	966	4	US-08-766-738-2 Sequence 2, Appl
30	34.8	2.2	1608	4	US-08-622-166A-1 Sequence 1, Appl
31	34.8	2.2	1608	4	US-08-622-166A-1 Sequence 1, Appl
32	34.6	2.2	1621	4	US-08-933-750C-96 Sequence 96, Appl
33	34.4	2.2	3387	1	US-07-867-106-2 Sequence 2, Appl
34	34.2	2.2	3387	5	US-08-261-822A-5 Sequence 5, Appl
35	34.2	2.1	3215	2	PCR-US95-07744A-5 Sequence 5, Appl
36	33.8	2.1	3215	2	US-08-426-627-1 Sequence 1, Appl
37	33.4	2.1	4817	1	US-07-951-715A-18 Sequence 18, Appl

ALIGNMENTS

38	33.4	2.1	3466	2	US-08-468-036-38 Sequence 38, Appl
39	33.4	2.1	3466	3	US-08-376-843-38 Sequence 38, Appl
C 40	33.4	2.1	4817	1	US-08-459-448A-18 Sequence 18, Appl
C 41	33.2	2.1	2188	1	US-07-865-662F-10 Sequence 10, Appl
42	33.2	2.1	1253	3	US-08-786-606-6 Sequence 6, Appl
43	33	2.1	4766	5	PCR-US93-07261-10 Sequence 10, Appl
44	32.8	2.1	1498	1	US-08-118-469A-1 Sequence 1, Appl
45	32.8	2.1	1498	3	US-08-909-119-1 Sequence 1, Appl

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION
; APPLICANT: DOMER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-6739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)583-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpc F15
; US-08-232-463-14

Query Match 4.8%; Score 75.8; DB 2; Length 7218;
Best Local Similarity 5.6%; Pred. No. 2.1e-11;
Matches 23; Conservative 237; Mismatches 149; Indels 0; Gaps 0;

QY 999 gaaatgcccaggtatcatattgtgtactataagccagaaatgtgatggat 1058
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1467 GCAAGTATTAAAGACATAGACAAATTGTGACRRRRRRRRRRRRRRRR 1408
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1059 ggatatgtgttctgtacttaccagaagagtgagtttgcctgaagcatatccg 1118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

ORIGINAL SOURCE

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 3253 base pairs

```

MOLECULE TYPE: DNA (genomic)

/ MATCORN SOURCE: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
;

```

;      HYPOTHETICAL: NO
;      ANTI-SENSE: NO
;      ORIGINAL SOURCE:
;      ORGANISM: Plasmodium vivax
;

```


US-08-676-974-2

Query Match 2.4%; Score 38.2; DB 3; Length 2277;

Best Local Similarity 26.0%; Pred. No. 0.25; Indels 0; Gaps 0;

Matches 133; Conservative 79; Mismatches 299;

QY 900 aaattgtcgaagatcccaagatccgggagatctcaacatcatgatgatctt 959
 DB 313 AAAAAAGCAAAAGTNGCNGATYARAARAGCNGMNTNATHTMGNAAYITNMSNTTYAAR 372
 QY 960 cctaccaagatgactgtatatttgcctccaagtagaataagtcaggattac 1019
 DB 373 TGWMSGARGAAYYTNARACNGTNTTYGNCARTTYGNGCNGTNTNGARGTNAY 432
 QY 1020 atattgttactataaggcagaatgatgatgataggatggttcttact 1079
 DB 433 ATHCCNMGNAARCCNGAYGNNARATGMNGNTTYGNTTYGTCNCARTTYAARAAYTN 492
 QY 1080 tacacaagaagtgcagtttgcctgaaagcatataccgagttgcaaacgcagctca 1139
 DB 493 YTNGARCGCNGNARACNTNARAGNATGAYATGAARATTHAARGMNGNCGTN 552
 QY 1140 aaagtgtgagcgtgatttcaacacatcataaaacagacacatacgtgcccgaacct 1199
 DB 553 GCGTNGAYTGGCNGTNGCNAARGAAYARAYAAAGAYAACNARMSNGTNGMNGNATH 612
 QY 1200 cccctgttgagaggttggaagaagaagtgcgaagaagaagcagtcataaaga 1259
 DB 613 GNGARGARARARMSNCAYGARMSNARCAVCARGARMSNGTNAARARARAGMNGNGAR 672
 QY 1260 gttagagagatagaagaagaagtgcgaagaagtgcgaagaagtcacctattcagt 1319
 DB 673 GAGAGAGATVTCARARARARARARAYGAYGATGATGATGATGATGATGATGATGAT 732
 QY 1320 aaactgttgaaggttttaagaagcctccacagagatgaagaagactcttaagaagcgtg 1379
 DB 733 GNGTNTTYGAYGAYGARAYGARAYGARAYGARAYGARAYGARAYGARAYGARAYGARAY 792
 QY 1380 agcaagaagaagtgcagtttgcagtgac 1410
 DB 793 GTNCARATHCARARARMSNGCNGTNAARMGNC 823

RESULT 8
 US-09-098-487-2
 ; Sequence 2, Application US/09098487
 ; Patent No. 5917025

; GENERAL INFORMATION:
 ; APPLICANT: COLLINS, Kathleen
 ; TITLE OF INVENTION: Human Telomerase
 ; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Science & Technology Law Group
 ; STREET: 268 Bush Street, Suite 3200
 ; CITY: San Francisco
 ; STATE: CA

; COUNTRY: USA
 ; ZIP: 94104

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/098,487

; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Osman Ph.D., Richard A

; REGISTRATION NUMBER: 36,627
 ; REFERENCE/DOCKET NUMBER: UCB96-055
 ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415)343-4341
 ; TELEFAX: (415)343-4342
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2277 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-09-098-487-2

Query Match 2.4%; Score 38.2; DB 4; Length 2277;

Best Local Similarity 26.0%; Pred. No. 0.25; Indels 0; Gaps 0;

Matches 133; Conservative 79; Mismatches 299;

QY 900 aaattgtcgaagatcccaagatccgggagatctcaacatcatgatgatctt 959
 DB 313 AAAAAAGCAAAAGTNGCNGATYARAARAGCNGMNTNATHTMGNAAYITNMSNTTYAAR 372
 QY 960 cctaccaagatgactgtatatttgcctccaagtagaataagtcaggattac 1019
 DB 373 TGWMSGARGAAYYTNARACNGTNTTYGNCARTTYGNGCNGTNTNGARGTNAY 432
 QY 1020 atattgttactataaggcagaatgatgatgataggatggttcttact 1079
 DB 433 ATHCCNMGNAARCCNGAYGNNARATGMNGNTTYGNTTYGTCNCARTTYAARAAYTN 492
 QY 1080 tacacaagaagtgcagtttgcctgaaagcatataccgagttgcaaacgcagctca 1139
 DB 493 YTNGARCGCNGNARACNTNARAGNATGAYATGAARATTHAARGMNGNCGTN 552
 QY 1140 aaagtgtgagcgtgatttcaacacatcataaaacagacacatacgtgcccgaacct 1199
 DB 553 GCGTNGAYTGGCNGTNGCNAARGAAYARAYAAAGAYAACNARMSNGTNGMNGNATH 612
 QY 1200 cccctgttgagaggttggaagaagaagtgcgaagaagaagcagtcataaaga 1259
 DB 613 GNGARGARARARMSNCAYGARMSNARCAVCARGARMSNGTNAARARARAGMNGNGAR 672
 QY 1260 gttagagagatagaagaagaagtgcgaagaagtgcgaagaagtcacctattcagt 1319
 DB 673 GAGAGAGATVTCARARARARARARAYGAYGATGATGATGATGATGATGATGATGAT 732
 QY 1320 aaactgttgaaggttttaagaagcctccacagagatgaagaagactcttaagaagcgtg 1379
 DB 733 GNGTNTTYGAYGAYGARAYGARAYGARAYGARAYGARAYGARAYGARAYGARAYGARAY 792
 QY 1380 agcaagaagaagtgcagtttgcagtgac 1410
 DB 793 GTNCARATHCARARARMSNGCNGTNAARMGNC 823

RESULT 9
 US-08-466-603-1
 ; Sequence 1, Application US/08466603
 ; Patent No. 5726018

; GENERAL INFORMATION:
 ; APPLICANT: Pasternack, Gary R.

; TITLE OF INVENTION: No. 5726018el Mammalian Protein Associated With
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner, Birch, McKie & Beckett
 ; STREET: 1001 C Street, N.W.
 ; CITY: Washington, D.C.
 ; STATE: District of Columbia
 ; COUNTRY: U.S.A.

; ZIP: 20001
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

Page 6

```

1 SOFTWARE: PatentIn Release #1.0, Version #1.22
2
3 CURRENT APPLICATION DATA
4 APPLICATION NUMBER: US/08/314,503A
5
6 FILING DATE:
7
8 CLASSIFICATION: 435
9
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Fosnotske Esq., Laurence H.
12 REGISTRATION NUMBER: 34,698
13 REFERENCE/DOCKET NUMBER: 1107,47218
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: 202 508-9153
16 TELEFAX: 202 508-9299
17
18 INFORMATION FOR SEQ ID NO: 1:
19 SEQUENCE CHARACTERISTICS:
20
21 LENGTH: 1052 base pairs
22
23 TYPE: nucleic acid
24
25 STRANDEDNESS: double
26
27 TOPOLOGY: linear
28
29 MOLECULE TYPE: cDNA
30 ORIGINAL SOURCE:
31 ORGANISM: Homo sapiens
32
33 FEATURE:
34
35 NAME/KEY: CDS
36 LOCATION: 97..843
37
38 OS-08-314-503A-1

```

LOCATION: 97..843
US-08-314-503A-1

QY	1260	gttggagagataataaagaagaatagaggaggtggaagataaagaagtaaccttaacgt	1319
Db	598	GATGATGAGAGGAGGAGATGAGCATAGCATAGAGGAGGATGATGAAGAAGATGCTCAGGTACTGGAA	657
QY	1320	aaactcttggaaagttttaaagagctccaacgagatgaagagaacttctaagagagctg	1379
Db	658	GACGAGAGAGACGACGAGATGAGGAGGAGGAGAAAGTGAAGAGGAGACGTGATGAGAAAGAG	717
QY	1380	agcaagaagaagaatgtagcttcttgatggaacttaaatggaagcaactgaggtagaataa	1439
Db	718	GAGGAGAGATGAAGAAAGCTTATACGATGAGAGAGATGATGACGAGGAAGATGAAGAAAG	777
QY	1440	ctttttg	1446
Db	778	CTTGGTG	784

RESULT 11
 US-08-468-066-1
 : Sequence 1, Application US/08469066
 : Patent No. 5756676
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Pasternack, Gary R.
 :
 : TITLE OF INVENTION: No. 5756676el Mammalian Protein Associated with
 : TITLE OF INVENTION: Uncontrolled Cell Division
 :
 : NUMBER OF SEQUENCES: 9
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: Banner, Birch, McKie & Beckett
 :
 : STREET: 1001 G Street, N.W.
 :
 : CITY: Washington, D.C.
 :
 : STATE: District of Columbia
 :
 : COUNTRY: U.S.A.
 :
 : ZIP: 20001
 :
 : COMPUTER READABLE FORM:
 :
 : MEDIUM TYPE: Floppy disk
 :
 : COMPUTER: IBM PC compatible
 :
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 :
 : SOFTWARE: Patent In Release #1.0, Version #1.25
 :
 : CURRENT APPLICATION DATA:
 :
 : APPLICATION NUMBER: US/08/468, 066
 :
 : FILING DATE: 06-JUN-1995

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,503
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Posorske Esq., Laurence H.
REGISTRATION NUMBER: 34,658
REFERENCE/DOCKET NUMBER: 1107.47218
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 508-9153
TELEFAX: 202 508-9299
INFORMATION FOR SEC ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1052 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 97..843
US-08-466-717-1

```

Query Match	2.3%;	Score 36.6;	DB 4;	Length 1052;
Best Local Similarity	49.7%;	Pred. No. 0.48;		
Matches	93;	Conservative	0;	Mismatches 94;
				Indels 0;
				Gaps 0;

QY 1260 gttgaggaatagaaanaaactaaagaaggttgagaaataaagaagtcacatttcagt 1319
Db 598 GATGATGAGAGAGAGAGATGAGATGAGAGAGAGATATGATGAAGTCTCTCAGGTAGTGGAA 657
QY 1320 aaactgtttgaaggttttaaaagctcccaacagatgaagaaacttctaagaagctg 1379
Db 658 GAGCAGAGAGAGCAGGAGATGAGAGAGAGAGAGAGTGAAGGAGCGTGACTGGAGAGAG 717
QY 1380 agcaagaagaatgtgagtgttttgatgtgaccttaaatgtgaagcaactgaggtagaanaa 1439
Db 718 GAGCAGAGATGACGAAGGTATTAACCATGAGAGAGTACGTACGACGAGGAAGATGAAGAAGAG 777
QY 1440 ctttttg 1446
Db 778 CTTGTGTG 784

RESULT 13
 PCT-US95-12414-1
 Sequence 1 Application PC/TUS9512414
 GENERAL INFORMATION:
 APPLICANT: Pasternack, Gary R.
 APPLICANT: Kuhnjoa, Francis P.
 TITLE OF INVENTION: Novel Mammalian Protein Associated With
 TITLE OF INVENTION: Uncontrolled Cell Division
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Allegretti, Ltd.
 STREET: 1001 G Street, N.W.
 City: Washington, D.C.
 STATE: District of Columbia
 COUNTRY: U.S.A.
 ZIP: 20001
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/12414
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,503

•
•
•
•

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 1999, 00:46:49 ; Search time 24.66 Seconds

(Without alignments)
654.388 Million cell updates/sec

Title: US-09-075-375-4

Perfect score: 7495
Sequence: 1 MALSLHTVFLCKEALNLNLA.....MEASEVKELGKALPIKVR 1412

Scoring table: BIOSUM62

Searched: 119832 seqs, 11428610 residues

Database: Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/PCITUS9.COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/Backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147	2.0	2482	1	US-08-328-254-6
2	135.5	1.8	1312	2	US-08-592-126-148
3	132	1.8	776	1	US-08-021-601-2
4	132	1.8	776	1	US-08-082-849B-2
5	132	1.8	776	3	PCR-US94-01624-2
6	124	1.7	1581	2	US-08-404-531B-6
7	124	1.7	1498	2	US-08-404-531B-28
8	115.5	1.5	1130	2	US-08-460-309-2
9	115.5	1.5	1130	2	US-08-460-309-4
10	115.5	1.5	1130	2	US-08-125-077-2
11	115.5	1.5	1131	2	US-08-125-077-4
12	115	1.5	1435	2	US-08-568-459A-4
13	112.5	1.5	3248	3	US-08-353-700-1
14	112.5	1.5	3248	3	PCR-US95-16216-1
15	111.5	1.5	3418	2	US-08-603-753D-4
16	109	1.5	3135	1	US-08-323-170B-2
17	109	1.5	425	2	US-08-986-963-2
18	108.5	1.4	3418	2	US-08-639-501-2
19	104.5	1.4	984	2	US-08-242-932-2
20	104.5	1.4	984	3	US-08-714-481-2
21	104.5	1.4	984	3	PCR-US95-06111-2
22	103.5	1.4	837	2	US-08-426-627-23
23	103.5	1.4	450	2	US-08-712-939B-1
24	102	1.4	666	2	US-08-441-139-16
25	101.5	1.4	650	1	US-08-325-071-56
26	101.5	1.4	688	1	US-08-325-071-57
27	101.5	1.4	984	2	US-08-257-073-3
28	101.5	1.4	984	2	US-08-184-009-120
29	101.5	1.4	3075	2	US-08-460-309-5
30	101.5	1.4	3075	2	US-08-125-077-5
31	101.5	1.4	984	2	US-08-458-356-120
32	101.5	1.4	322	2	US-08-904-031-1
33	100.5	1.3	528	1	US-08-160-670A-5
34	100.5	1.3	789	2	US-08-471-033-32
35	100.5	1.3	789	2	US-08-471-044-32
36	100.5	1.3	789	2	US-08-463-483A-32
37	100.5	1.3	789	2	US-08-471-046A-32
38	100.5	1.3	789	2	US-08-470-566B-32
39	100.5	1.3	789	2	US-08-838-219B-4

40	100	1.3	1196	1	US-08-144-121-4	Sequence 4, Appl
41	100	1.3	836	2	US-08-426-627-6	Sequence 6, Appl
42	100	1.3	1196	2	US-08-735-893-4	Sequence 4, Appl
43	99.5	1.3	1209	3	PCR-US95-04589-107	Sequence 107, App
44	99	1.3	654	2	US-08-441-139-11	Sequence 11, Appl
45	98.5	1.3	1107	1	US-08-366-577-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1	US-08-328-254-6	Sequence 6, Application US/08328254	Patent No. 5710022
GENERAL INFORMATION:			
APPLICANT:	Zhu, Xueliang		
INVENTOR:	Lee, Wen-Hua		
TITLE OF INVENTION:	A No. 5710022e1 Nuclear Mitotic Phosphoprotein		
NUMBER OF SEQUENCES:	8		
CORRESPONDENCE ADDRESS:			
ADDRESSEE:	Campbell and Flores		
STREET:	4370 La Jolla Village Drive, Suite 700		
CITY:	San Diego		
STATE:	California		
COUNTRY:	USA		
ZIP:	92122		
COMPUTER READABLE FORM:			
MEDIUM TYPE:	Floppy disk		
COMPUTER:	IBM PC compatible		
OPERATING SYSTEM:	PC-DOS/MS-DOS		
SOFTWARE:	PatentIn Release #1.0, Version #1.25		
CURRENT APPLICATION DATA:			
APPLICATION NUMBER:	US/08/328,254		
FILING DATE:	24-OCT-1994		
CLASSIFICATION:	435		
PRIOR APPLICATION DATA:			
APPLICATION NUMBER:	US 08/141,239		
FILING DATE:	22-OCT-1993		
ATTORNEY/AGENT INFORMATION:			
NAME:	Campbell, Cathryn A.		
REGISTRATION NUMBER:	31,815		
REFERENCE/DOCKET NUMBER:	P-CJ 1191		
TELECOMMUNICATION INFORMATION:			
TELEPHONE:	(619) 535-9001		
TELEFAX:	(619) 535-8949		
INFORMATION FOR SEQ ID NO:	6:		
SEQUENCE CHARACTERISTICS:			
LENGTH:	2482 amino acids		
TYPE:	amino acid		
TOPOLOGY:	linear		
MOLECULE TYPE:	protein		
US-08-328-254-6			
Query Match	2.0% Score 147; DB 1; Length 2482;		
Best local similarity	18.4% Pred. No. 0.0034;		
Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;			
379 LVERLEKTAEGEKKILKEAVEIEEVE-----KEVE-KVRDIEMLFQRLLEGFKLQ 432			
315 LSETLSEKTKVSSILNKRREIELNENGTLEKINASINQERMDLQK-SESFANYID 373			
433 DEENVRRLS---KEKKIL-----			
374 EREKISLSLQKQKILLQCEBETGNAYEDLSQYKAAQENSKLDECLNECTSLCE 433			
450 --NELQEAATEVEKLEGRALP--IRKLRMALAPNSFLANHETIKKYVSKLPGH---- 501			
434 NKKNLE-----CKKAFKHEQHEFLTKLAFAEERNQNMLELETVQALSEMIDNQNS 489			
501 KRESGWEDYFGSIVAKICSSRIPIPRFKSPRI-----CCGLDSRGDLQLFSGHK 551			

```

Db 490 KSEAGLKEIIMTLKEQNKKQKQFVNDLQENQJMKVMTKHCCNLESEPIR----- 544
QY 552 HNLSPASINONVPGKSGCKFPKDVALLMWEKKQGFATAIYAFILSVASADAVDAI 611
Db 544 -----NSVKERESERNO-CNFKPOMDEVEKEISLDSYNALQVLEAMLNKELKIQEES 596
QY 612 KTCICL-----LKECRLEL-AKCISNP-----ACAN 637
Db 597 KECICLQHEIOTIRGDLSTNLODMQSEISLKDCEIDEEFYISGPHLETSQNDNAH 656
QY 638 VAC-LQICNNRPDETF--COIKGDLFENSVDENECASRRKCY--PRK--SDVGP- 690
Db 657 LQSLQITMKMLNELEKICEILQAEKYE--LVEIAND--SRSECIATARKMAEEVGKLL 711
QY 690 -----PVDPSPVLVQKFMKDFSGKWFTRGLNP-----TFAPFC 725
Db 712 NEVKILNDSSGLLHGEIVEDIPGGEF--GEQPNQHPVSLAPLDESNEYELTSLDKEV 768
QY 726 QLHEFHTENKLVGNLSWIRTPDGGFTRSAVQKFEVQPKYPIGLYNHNDNYLLYQD-- 784
Db 769 QMFHAELOEK-----FLSQSHKILHQQHCQ--MSKMSLQITYVDSL 810
QY 784 --DMYILSSKVENSPEDYIFVYKGRND-----AMDYIGGSVLYTRSAVL 826
Db 811 KANLVLTSLMIRNQGDLVKEMQGLEEGVPSLSSCVDPSSSLSLDSSFY--RALL 868
QY 827 PE-----SIPELOTAO-----KYGRDENTFIKTDNCGPEPL-----VERLEKVEE 873
Db 869 EQGDMSLSLNLEGNANSGSDEYFCSSLOJENLTRKETSAPKGYEELSECEYVR 928
QY 874 RTIIEVEEIEEVEER--VRDEKVTLFSEKREG-----FKELORDENELRELSKE 922
Db 929 QSL-----EKLSEKMEQOIKMKKEIOELPQLSSERQELDLKQYLSNEQMOQKLSV 984
QY 923 EMQVLDJLKWEATEVEKLF-----GRALPIRKIMAAVTH-----CPTS-- 961
Db 985 TLMESKLAEEKQTDLSLEVAARLQGLDLSRSILGIDTEDAIGRNESCDISKE 1044
QY 961 -----PCHDIRIFFSSDDIGRLGTRKINGTFLKILPFI-----QSADL-- 1003
Db 1045 HTSETTERPKHD-VHQICDKDAQOQDLNDIEKITETGAK--PTGEGSGQSPDITVE 1100
QY 1003 ---RTTGRS-----SRPISARSGFSKIGEDIVLPFRK----- 1034
Db 1101 PPGEDKTQSSSECSISELSESGPMALVPMDFLGNQEDIHNLQIRVKESTNENRLHLVIED 1160
QY 1034 -----NELKELTAPFLILKLVGLACAPLIVPSADAVDLKTCACILKRCRIELAKC 1084
Db 1161 RDRKVESLNMKEKELDLSKHLQVQMLT-----KTEAC-----TELKRI 1199
QY 1085 IANPACANVACLQTCNNRPDETEQIKCGDLFENSVDENECASRRKCYPRKSDIG- 1144
Db 1200 V-----GEL-----KX-----EMSDLSE 1212
QY 1144 --EFPAPDPSVLVONNISDFNGKWITTSGLMPTDAPCCQHEHTF-----GDNKLVGN 1197
Db 1213 KLEYSFCHQELLOVETSE-----GLNS-----DLEMHADKSSREDIGNAVAKVN 1258
QY 1198 ISMRITLDSGFFTSAAVQKFEVQDPNQPGLVYNHNDNEYH-----YODDWYILSSKIENK 1252
Db 1259 DSMKEHFLD-----VENELSRIRSEKASIEHEALYIHEADLEVYQTEKLCLEKDNENK 1310
QY 1253 P-----EDYIFVYVRGRANDAWGYGAAVYTRSSVLENSIIPLELKAASIGRDFSTFI 1306
Db 1311 QKVIYCLEBELSVVTSERMOUL--RGELDTMSKTTALDQISEKKEKTEQLESHQSQCL 1367
QY 1307 RPDNCGPEPALVERIEKVEGERIIVKVEYIEE--VEKEVEKGRIMLIFORLA 1363
Db 1368 HCTOYAAEAEVAKTELTQTLSSVSELLDKTHLOEKLOESKEQSOALSTRKCELENOIA 1427
QY 1364 EGFNELKQDENFVRE-----LSKEMEFLDEIK--MEASEVER 1400
Db 1428 ---OLNKEKELLVKESESLQARLESSEYKRLNVSKALFAALVER 1468

```

```

RESULT 2
US-08-592-126-148
; Sequence 148; Application US/08592126
; Patent No. 5821091
;
GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
;
US-08-592-126-148
Query Match 1.8%; Score 135.5; DB 2; Length 1312;
Best Local Similarity 19.7%; Pred. No. 0.012;
Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

```

Thu Oct 14 07:57:28 1999

us-09-075-375-4.rail

Page 3

```
QY      723 FDCOLHEHTEENKLVGILSWIRITPTDGGFFTRSAVOKFVODPPXPGLIYNHNEYLLQ  782
       | : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      669 FIPL-----IDNOSCCPYCVR-----POTAELOEVIDS-----              701

QY      783 DDWATLSSKYENSEPDYIFV--YKGRNDAMDYGGSVLTYRSKV-LPESIIPELOTAQ  839
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      701 -----LOSKLIAFDPKLTSESLKKRKRDEMDIGVPMROSIIIDKEKEIPELRNKLO  755

QY      840 KYVRDFNFETID-----NTCGPEPLVERLERKVEGERGTIIKE--  880
       | | | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      756 NVNRDIQR-LKNDIEBOTLLIGTIMPEESAKVCLLDTVIINERQMELKDYERKIAQA  814

QY      880 -----VEEIEEVKYVRDEVTLFSKLEFGFELODEENN-----LRELSKEE  923
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      815 KLOGIDLRIVQAVNOEKOEKOHLDTYSSKIELNRKLIODOCOFOILOHKSTTNELKSEK  874

QY      924 MDVLDTLKME-----ATEYEKLE  941
       : : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      875 LQISTNLDORROOLEEQYVELSTEVOSSLT  902
```

RESULT 3
US-08-021-601-2
; Sequence 2, Application US/08021601

```

1  GENERAL INFORMATION:
2  APPLICANT: Leppla, Stephen H.
3  APPLICANT: Klimpel, Kurt R.
4  APPLICANT: Nichols, Peter J.
5  APPLICANT: Aroia, Naveen
6  APPLICANT: Singh, Yogendra
7  TITLE OF INVENTION: ANTIDRAX TOXIN FUSION PROTEINS AND
8  TITLE OF INVENTION: RELATED METHODS
9  NUMBER OF SEQUENCES: 12
10 CORRESPONDENCE ADDRESSES:
11 ADDRESSEE: Needle & Rosenberg, P. C.
12 STREET: 133 Carnegie Way, Suite 400
13 CITY: Atlanta
14 STATE: Georgia
15 COUNTRY: USA
16 ZIP: 30303
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: PatentIn Release #1.0, Version #1.25
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/021,601
24 FILING DATE: 19930212
25 CLASSIFICATION: 514
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Spiralt, Gwendolyn D.
28 REGISTRATION NUMBER: 36,016
29 REFERENCE/DOCKET NUMBER: 1414,057
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 404/688-0770
32 TELEFAX: 404/688-9880
33 INFORMATION FOR SEQ ID NO: 2:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 776 amino acids
36 TYPE: AMINO ACID
37 TOPOLOGY: linear
38 MOLECULE TYPE: protein
39
40 US-08-021-601-2

```

Query Match	1.88;	Score 132;	DB 1;	Length 776;
Best Local Similarity	17.28;	Pred. No. 0.0098		
Matches 139;	Conservative 117;	Mismatches 254;	Indels 300;	Gaps 31.
QY	733	EENKLVGNLSWRITPDGCFRTSANOVKFQODPKYIGILYNHD---	NEVILLIYDDMY--	I 787
	+	+	+	+

Dd	72	EMKALIG---	KIYIDGOLITHISLEALSEDDKRIKIXIYGDALHBBHYAYKAGEPV	128
Qy	788	LSKVENSPED:TFVYKGRNDAMDYGSGSVLTYSANVPESILPELOTAOKRGRENT	847	
Dd	129	L---VIOSSDEYVENTEKALN-----	-VYIEGLIKILSRDLISKINOPQOKELDVLT	176
Qy	848	FIKTDNCG-----	-PEPLVSELEKVEEGERTIK-----	879
Dd	177	IKNASDSDGDDJLFTNOLKEHPTDSVEFLLEONSNENVOEVRKAFAYITEPQHNDVLOLY	236	
Qy	879	-----	-VVEIEEVEYKVRDEKVLTFSKLEGEKE-----	908
Dd	237	APFAFNMDKFNPOEILUSLELKQRMLSREKWEKIKQHYQWSDSLSEEGILTKL	296	
Qy	908	---LORDEENFLELSKEENDVIDGJLKMATEV---	-EKLSGRALPT-----	RKL 950
Dd	297	QIPEPKRDI1HSLSOEKELRLRQIDSSDFLSTEKEFLPKLQIDIRDSLSEEEKL	356	
Qy	951	MAVATGFTSPCHDRJEFESDDIGRLGTRIKRINGFLKILPPIOSAD---	LRTTG 1006	
Dd	337	L-----	-NRIOVDSNP-----	LSEK -EKELTKLKDIOPIYDINQRLQDTG 396
Qy	1007	GRSSRPL-----	-SAFRSGFSKGIFDIYPLPSKNELELAPLLKLVG	1049
Dd	397	GLIDSPSLINDVRKQYKRDIQIDALLHOSIGSTLYENKILYENNMINNLATL-----	451	
Qy	1050	VLAACFLIVSADAVDALTKACGLGCRLEIAKCIANPACANVACLOTCNNRPDETC	1109	
Dd	451	-----	-GADVDST-----	459
Qy	1110	QIKCGDLEFENSVD-EFNECAVSRKKCYPRK---	-SDLGEPPADPPSVLQNFNISDENG	1164
Dd	459	-----	-DNKIRNGJFNEFKKXFKSISNMIVDINERALDNERKMKWIOQSPOTR	510
Qy	1165	KWYIISGLNPTFDAFCOLHEFTEGDNKLV-	-GNISWRIKILDSGFTTRSAVQKFEVDP	1222
Dd	511	AGYLENG-----	-KLIIORNGLEIKBOV-----	534
Qy	1223	NQPGVLNHNDEYLAHYODMWILSKTEKNPEYIIVYRGRNDAMDYGCAVYTR--S	1280	
Dd	534	---IKOSEKEYIRI-DARVAPSKIDTIOQ---	-AOLINQEWKALGLPKTKLIT	584
Qy	1281	SVLPNSIIPLEKEAKSIGRDEFSTFIRTDQCPERALVERLEKTEVEEGE-RIIVKEV-	1338	
Dd	585	FNVHRKYSNIVEASAYLLIMKKNKNIOS-----	-LLKATNYLVDSNGRFAFTDITL	636
Qy	1338	-----	-BEIEEVEKEVEKVGRTMTLFLORLAEGFENELKODEENFVR-----	1380
Dd	637	PNIAEQYHOPEIIEOVHSGVLVPEBSRSLILHGPSKVY-ELFNDSEGFIEHFCAVVDY	695	
Qy	1380	-----	-LSKEEME-----	FLDEIKMESEV 1398
Dd	696	AGYLLDRKMSDLVTKSKKFDIFEEBSNLT	725	

RESULT 4
 US-08-082-849B-2
 : Sequence 2. Applicant us/08082849B
 : Patent No. 5677274
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: LePrel, Stephen H.
 : APPLICANT: Kimpel, Kurt R.
 : APPLICANT: Aroia, Naveen
 : APPLICANT: Singh, Yogendra
 : APPLICANT: Nichols, Peter J.
 : TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
 : TITLE OF INVENTION: Related Methods
 : NUMBER OF SEQUENCES: 35
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: Townsend and Townsend and Crew LLP
 : STREET: Two Embarcadero Center, Eighth Floor
 : CITY: San Francisco
 : STATE: California

COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/082,849B
 FILING DATE: 25-JUN-1993
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/021,601
 FILING DATE: 12-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Weber, Kenneth A.
 REGISTRATION NUMBER: 31,677
 REFERENCE/DOCKET NUMBER: 15280-161-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0300
 TELEFAX: (415) 576-0200
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 776 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-082-849B-2

Query Match 1.8%; Score 132; DB 1; Length 776;
 Best Local Similarity 17.2%; Pred. No. 0.0098;
 Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 ENKLVGNISWIRTPDGGFTFRSAVQKFPYVPGILYVND--NEYLLYODDWY--I 787
 DB 72 EYKKAIGG--KIYVDDITKHISLEALSDKKIKIDYKDALHHEHYVAKGEYEPV 128
 QY 788 LSKVENSPEDEIFVYGRNDAMDYGGSVLYTFSAVLPESIIPELOTAOKGRDENT 847
 DB 129 L--VIQSSSEYVENTEALN-----YYEIGKILSDITLSKINQPYOKFLDVLT 176
 QY 848 FIKDNTNG-----PEPPLVERLKKVEGERITIK----- 879
 DB 177 IKNASDSGODLLFTNQLEKHEPTDFSEVFLPQNSNEVQEVAKAFAYVLEPHQHDVLY 236
 QY 879 -----EVEIEEVEKVRKDEYTLFSKLEFGK----- 908
 DB 237 APEAFNMYDKFNEQETINLSLELKDORMLSRYEKWEKIKOHYOHMSDSLSEGRGLIKL 296
 QY 908 --LQREDENFLRELKREMDVLDLKEATEV---EKLFGRALPI-----RKL 950
 DB 297 QIPTRKDDIHSLSQEKELKRIQIDSSDLSTEKEKELKQIDIRSLSEEEKEL 356
 QY 951 MAVATHTCTSCCHDIRIRFFSSDDGIGRLGIRKRIRNGTFLIKLIPQOSAD---LRTTG 1006
 DB 357 L-----NRIOYDSSNP-----LSEK--EKEFFLKKLIDQIPDINQRLQDTG 396
 QY 1007 GRSRPL-----SAFRSGFSKGFIDVLPSPSKNELKELTAPLLKLVG 1049
 DB 397 GLIDSPSINLDVROKQYKRDIONITALLHQSIGSTLYNKIYIYENMNINNTLATL----- 451
 QY 1050 VLACAFILVPSADAVDALTKTACLLKGCRIELAKCIANPACANAVACLOICNNRDETEC 1109
 DB 451 -----GADLVDS----- 459
 QY 1110 QIKGDLFENSVDE--FNECAVSRKKCVPRK---SDIGEPADPSPVLVONENISDFNG 1164
 DB 459 -----DNTRKIRGAFNEFKKNFKYSISSNYMIVDINERPALDNRLKWRIDLSPTDR 510
 QY 1165 KMYITSGINPTFDFACQLHFFHTEGDKLV--GNISMRKITLDSGFTTSAVQKFPYQDP 1222
 DB 511 AGYLENG-----KILQNRNIGLEIKQV----- 534

QY 1223 NOGVLYNHDEYLYHODDWYILSSKIENKPEDYIFVYGRNDAMDYGGAAYVYTR--S 1280
 DB 534 ----LIKQSEKEYIRI-DKAVPKSKIDRKIQE-----ADLNINQENKALGIPKTKLIT 584
 QY 1281 SVLPNSTIIPLEKAASIGRDSFTIRTDNTGCPGEPALVERIEKTEVEGE-RIVKEV-- 1338
 DB 585 FVNHRYASNIYSAVLLINEMKRNIOQD-----LIKVTNLYLDGNGRAFFFTDITL 636
 QY 1338 -----EIEIEEVEKVEKVGRTENTLFORLAEGFNEIKQDEENFVRE----- 1380
 DB 637 PNIAEQTHODELYEQVHSGLYPESRSILHGSKGV-ELRNDSEGIHEFGHAVDY 695
 QY 1380 ----LSKEME-----FLDEIKMEASEV 1398
 DB 696 AGYLLDKNOSDLVTNSKRFIDIFKEGSNL 725

RESULT 5
 PCT-US94-01624-2
 Sequence 2, Application PC/TUS9401624
 GENERAL INFORMATION:
 APPLICANT: Leppla, Stephen H.
 APPLICANT: Klimpel, Kurt R.
 APPLICANT: Ariora, Naveen
 APPLICANT: Nichols, Yogendra
 APPLICANT: Singh, Peter J.
 TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
 STREET: Stewart Street Tower, 20th Floor, One Market
 STREET: Plaza
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94105

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/01624
 FILING DATE: June 25, 1993
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Weber, Kenneth A.
 REGISTRATION NUMBER: 31,677
 REFERENCE/DOCKET NUMBER: 15280-115
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 543-9600
 TELEFAX: (415) 543-5043
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 776 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-01624-2

Query Match 1.8%; Score 132; DB 3; Length 776;
 Best Local Similarity 17.2%; Pred. No. 0.0098;
 Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 ENKLVGNISWIRTPDGGFTFRSAVQKFPYVPGILYVND--NEYLLYODDWY--I 787
 DB 72 EYKKAIGG--KIYVDDITKHISLEALSDKKIKIDYKDALHHEHYVAKGEYEPV 128
 QY 788 LSKVENSPEDEIFVYGRNDAMDYGGSVLYTFSAVLPESIIPELOTAOKGRDENT 847

```

Db 129 L---VIOSEEDYVENTEKALN-----VYEIGKILSRDLISKINOPYOKFLDVINT 176
QY 848 FIKDNTNG-----PEPPLVERLEKKEVEEGRTIIR-----879
Db 177 IKMNSDSDGDLFTNOLKEHPTDSEVEELONSNEVOEFAKAFAYIEPOHNDVLOLY 236
QY 879 -----EVEIEEVEKVRDKEVTLFSKLEGEFKE-----908
Db 237 APEANMYDKTNEQININLSELEKDLQRLSLRYEKMEKIKQHYOHMSDSLSEGGILKLL 296
QY 908 ---LORDEENFLRELKEMOVLGLKMEATEV---EKLFGRLPT-----RKL 950
Db 297 QIPIEPKDDIHSLSQEEKELKRIQIDSSDFLSTEKEPELKLQIDIRDSLSEEEKEL 356
QY 951 MAVATHCTSPCHODIRFRESDDGIGRLGTRKRINGTFLKILPIPOSAD---LRTTG 1006
Db 357 L-----NRIOVDSSNP-----LSEK--EKFELKLLKDIOPIYINORLODTG 396
QY 1007 GRSSRPL-----SAFRSGFSKGIPIVPLPSKNELEKLTAPLLKLVG 1049
Db 397 GLIDSPSINLDRKQYKRDIONIDLHQSIGSTLYNKIYIYEMNNINNLATL-----451
QY 1050 VLACAFLLVPSADAVDAKTCACLLKGCRIELACIANPACANAVACLOTGNNRPDETEC 1109
Db 451 -----GADLVDS*-----459
QY 1110 QIKGDLFENSVDYD--FNECAVSRKRCVPRK--SDLGEFPAPPVSVLVQNFNISPENG 1164
Db 459 -----DNTKINGITNEERKKNKFISSNYMIVDINERPALDNEKLKRIQISPTR 510
QY 1165 KWIYITSGINPTFDFACOLHEFHTEGDNKLV--GNISWIRIKTLDSGFETRSAVOKFVODP 1222
Db 511 AGYLENG-----KLIIQRNIGLEIKDVO-----534
QY 1223 NOPGVLYNHNDYLYHXYODDWYLLSKINKPEPDIYFVYVYRGNDAMDYGAVYTR--S 1280
Db 534 ---IIOSEKEXIRI--DAKVPRKSIDTKIOE---AQLNINEMWNALELPKTKLIT 584
QY 1281 SVLPMSIIPELKAKSIGRDFSTFRTDNCGPEPALVERIEKTEVEGE--RIYKVEV-- 1338
Db 585 FVNHRYKSNYVESAYLLINEMKNNIQSD-----LIKATYNTLVGNGRFVETDITL 636
QY 1338 -----EEIEEVEKEVEKVRTEMTLFLQRLAEGFENELKODENFVE-----1380
Db 637 PNIAOYTHODIYEQVHSKGLYVPSRSILHGPSKGV--ELRNSSEGIHEFGHAVDY 695
QY 1380 ---LSKEME-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNOSDLVTSKRFIDIFKEEGSNL 725

```

RESULT 6
US-08-404-531B-6
Sequence 6, Application US/08404531B
Patent No. 5863724

GENERAL INFORMATION:
APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
Patent No. 5863724
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESS: One Liberty Place 46th. Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/404,531B
FILING DATE: 15-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Beardsell, Lori Y.
REGISTRATION NUMBER: 34,293
REFERENCE/DOCKET NUMBER: BYLR-0003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1581 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-404-531B-6

```

Query Match 1.7% Score 124; DB 2; Length 1581;
Best Local Similarity 18.8%; Pred. No. 0.15; Matches 221; Indels 178; Gaps 26;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

```

QY 868 KVEGERTIIRKVEIEEVEKVRDKEVTLFSKLEGEFKE--ELORDEENFLRELKREND 925
Db 480 KLSQAQRITL-----EYSNERLKQT--NEMLRGILLKLYAMENFCSRVKETRKR 528
QY 926 VLGLKMEA--TEVEKLFGRALPIRKLAHAVATHCTSPCHDIRFPSSD-----GIGLU 978
Db 529 EMTSLRAFAVYISISIFMTAIPDAVAILTFV-----GHVSEFKESDFSPVAFSL 580
QY 979 GTRKRKINGTFL-----KILPPIO-----SADLRTTGGRSSRPLSAFRSGFSKG 1023
Db 581 SLPHILVTLPLFLSSVYRSVRYKALVSQKLSFSLASAEIREOCAPREAPQGA-----G 636
QY 1024 IPDIVPL-----PSKNELEKLTAPLLKLVLGVLACAFLLVPSADAVDAKTCACLLKG 1076
Db 637 KYQAVPLKAVNRRKPAAREVRDLGFLQ-----RLPSTDG--DADNFCVQIIG 684
QY 1077 C-----RIELACIANPACANAVACLOTGNNRPDETEQIKGDLF 1117
Db 685 FFWTPDGIPTLSNITIRIPRGOLTMIVGVGGKSSLLATLG-----EMQKVGAVF 738
QY 1118 ENSVYDNEFCVSRKRCVPRKSDLGEFPAPDP-----SVLVQNFNI--SDF 1162
Db 739 WNSLPPD-----SEGRROQPRAGDSGRFGQEQORPCGYASQKWLINATIVENITRESF 793
QY 1163 NGKWY---ITSGINPTFDFACOLHEFHTEGDNKLVGNISWIRKT-----1205
Db 794 NKQRYKMWIEA--SLQPIDIL-----PHGQTOIGERGINTLSTGGORPDQCRPEPST 845
QY 1205 -----LDSGF-----FTRSAVOKFVODDPQOPGVLYNHNDYLYHXYODDWYLLSS 1247
Db 846 STPWIVFLDDPFSALDVHLSDLHMOAGILELLRDKDRVAVLYTHKLOYPHA--DWIIMAK 904
QY 1248 KIENKPEPDIYFVYVYRGNDAMDYGAVYTRRSVLPNSIIPLEKAKSIGRDFSTFTR 1307
Db 905 DGIQREGTAKTPQRSQQLFEHM-----KILMNRODQLEK-----ETVME 946
QY 1308 TDNTGPEPALVERIEKTEVEGERIIVKEVEIEEVEKE 1347
Db 947 RK---APEPS---VILPRAMSSRDGLLDEDEEIEEAESE 981

```

RESULT 7
US-08-404-531B-28
Sequence 28, Application US/08404531B
Patent No. 5863724
GENERAL INFORMATION:
APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
APPLICANT: Thomas, Gilbert Cote, and Robert Gagel

TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
 Patent No. 5863724
 NUMBER OF SEQUENCES: 49
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 ADDRESSEE: No. 5863724ris
 STREET: One Liberty Place 46th. Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/404,531B
 FILING DATE: 15-MAR-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Beardell, Lori Y.
 REGISTRATION NUMBER: 34,293
 REFERENCE/DOCKET NUMBER: BYLR-0003
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-568-3439
 TELEFAX: 215-568-3100
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1498 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-404-531B-28

Query Match 1.7%; Score 124; DB 2; Length 1498;
 Best Local Similarity 18.8%; Pred. No. 0.14;

Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

QY 868 KVEEGERTIIEVEIEEVEKAVDKETVLSKFEFGK--ELORDENELRELSKEMD 925
 DB 480 KLSQKQRTLL-----EYSENERIKQ--NEMLRGKLKLKLYAMENICSRVEKTRRK 528
 QY 926 VLDSIKMEA--TEYEKLFGRALPIRKIMAAVATHCFTSPCHDIRIFSSDD----GIGRL 978
 DB 529 EMTSLRAVAVTSISIFMNTAIPAAVLITV-----GHVSFFESPSFSVATASL 560
 QY 979 GITRKRIINGTFL-----KILPPIQ-----SADLRTGGSSRPLSAFSGFSKG 1023
 DB 581 SLFHILVPLFLSSVYRKAIVSVOKLSEFLSSAEIREQCAPREPAPOGA----G 636
 QY 1024 IFDIVPL-----PSKNEKELTAPLLIKVYLACAFILVPSADAVDLKTCACILKG 1076
 DB 637 KYQAVPLKVVARRKRAEVDLLGPLD-----RLTSTDS--DADNFCVOYIGG 684
 QY 1077 C-----RIELAKCIANPACANVACIQTCCNNRPDETECOIKGCDLF 1117
 DB 685 FFTWPDGIPLTSLNITIRIPRGQLTWVGQVCGCKSSLATLIG-----EMQKVSQAVF 738
 QY 1118 ENSVYDFENEGAVSRKCVPRKSDLGEPAPDP-----SVLVQNFNT--SDF 1162
 DB 739 WNSLPD-----SEGRRQOOPRAGSGRFGCOEORPCGYASQKPMWLNAVTEENITFESP 793
 QY 1163 NGKVV-----ITSGNPTFFDQDLHEFHTEGDKLIVGNSWIRKT----- 1205
 DB 794 NKQYKXKAVIENGSLQPIDIL-----PHGDTQIGEGINLSTGGQAPDOCREPST 845
 QY 1205 -----LDSGF-----FTRSAVQKFOVDPNQGVLYNHNDVEYLYODWYTLSS 1247
 DB 846 STPMIVFLDDPFSALDVHLSGHLMAQAGILRLDRDKRTVVLVTHKQYLPFA--DWIILAK 904
 QY 1248 KIENKPEDIYLFVYIRGNDAWDGIGGAVVYTRSSVLPNSIIPELERAANSIGRDESTFTR 1307

DB 905 DGTIOREGTLKFORSECOLFEHW-----KILMNQODELER-----ETVME 946
 QY 1308 TDNCGEPALVERIEKTEVEGERIIVKEIEEIEEVEKE 1347
 DB 947 RK---APEPS--QGLPRAMSRDGLUDEDEDEEBAESE 981

RESULT 8

US-08-460-309-2
 Sequence 2, Application US/08460309
 Patent No. 5837486
 GENERAL INFORMATION:
 APPLICANT: Engvall, Eva
 APPLICANT: Leivo, Ilmo
 TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
 TITLE OF INVENTION: Fragments and Uses Thereof
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,309
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/125,077
 FILING DATE: 22-SEP-1993
 APPLICATION NUMBER: US PCT/US 94/10730
 FILING DATE: 21-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/472,319
 FILING DATE: 30-JAN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/919,951
 FILING DATE: 27-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LA 9721
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1130 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-460-309-2

Query Match 1.5%; Score 115.5; DB 2; Length 1130;
 Best Local Similarity 20.0%; Pred. No. 0.46;

Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;

QY 294 NEFLAYODWYTLSSQLENKRPDIYFYVYRGNDAWDGYSVLYTRSPLE-SIIPN- 352
 DB 6 NDVKNEDHLNGKTRLENA-----DARNGDLRLTINDLGLKSLAIPND 49
 QY 352 ----LQRAKAVGVDENNFTITDNSCGPEPPLVRLKTEAEGEKLIIKAVAEIEEVEK 407
 DB 50 TAARKQ--AVKDKARQAND--TAKDVLAQITELHQNLDGLKKNYKLT--ADSVAKTAVVK 104

Thu Oct 14 07:57:28 1999

us-09-075-375-4.raii

Page 7

[illegible]

```

      FILING DATE:
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/125,077
      FILING DATE: 22-SEP-1993
      APPLICATION NUMBER: US PCT/US 94/10730
      FILING DATE: 21-SEP-1994
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/472,319
      FILING DATE: 30-JAN-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/919,951
      FILING DATE: 27-JUL-1992
      ATTORNEY/AGENT INFORMATION:
      NAME: Campbell, Cathryn A.
      REGISTRATION NUMBER: 31,815
      REFERENCE/DOCKET NUMBER: P-LA 9721
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (519) 535-9001
      TELEFAX: (619) 535-8949
      INFORMATION FOR SBO ID NO: 4:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 3111 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      US-08-460-309-4

Query Match          1.5%; Score 115.5; Db 2; Length 3111;
Best Local Similarity 20.0%; Pred. No. 2.4;
Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps

QY 294 NEFLHYDDWYLISQIQENKFPDYIFVYIRGRNDAMDGIGGSVIITRSPTLPE-SILPN- 352
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1987 NDVENEDHDJNJKTRTREN-----DARNGDLLRLINDLTGLKLSAIPND 2030
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 352 ----LQRAAKSVGDFNNFTTTNSCGEPPLVERLEKTAEGEKILLIKENAEIEEVEEK 407
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2031 TAARLO-AVKDKRAQAND--TAKDVLAQTTELHQNLIDGLKKNYNK--ADSVAKTNNAVK 2085
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 408 EVER---VRDTFML-----FORLGEFEKLQODENEFVRELSKEEELNELOMEATE 458
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2086 DPSKNKIILADDAIVAVKVLDEADRLIDKLPIKELEDNLKKNIS-EIKELINQAARKONS 2144
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 459 VE---KLFGRALPIRLRLMALAPHSNFLANHET----IKYYVG-----KLPGH 500
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2145 IKVSVSSGGDCIRTKPELIKSGSYNNIVNVKTAVDANLFLYLSAKFTDLAIEMRRGK 2204
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 501 KRFSNGW-----EDYGSIYVAKICGSRRIPIRFKRSPRICGDISRGIO 545
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2205 VSFIMWDVSGGVAREYEDLTIIDSYWRYIASRTGNRTI-----SVR--ALDGPKAS 2255
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 546 LESHGKHNLSPAHSINQNVPKNSGCKFKPDVALMWEMKGQFAKTAIVAIFILSVASKA 605
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2256 IVPSTHHSTP-----PQYTILDY-----DANMLLVGGILGTKL 2285
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 606 DAVALKTCTCLKECRYLEAKCI-----SNPACANYACLOTCNNRPDETCCOIGCD 659
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2290 KKADAVRVIT-----FTGCGETFYDNKP-----IDLWFRKEEGDCK----- 2328
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 660 LFENSVDVEFEKAVSKKCVPRKSD-----VGDPFV--PDPSSLVQKF-- 702
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2338 -----GCTVS-----POVESEGIOLFDEGEGIALVSRPIRWYNISTWMEKFFT 2377
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 702 -----DMKDF-----SGKW-----FI 712
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2372 FSSSLMLMYLTLRQLRPFMSVELTDGHIKSYLDLGSMASVYSNQNHNGKKKSTLSLI 2433
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 713 TRGLNPTEDAFDLCOLHEFHTEENKLV---GNLSWRIRPRD---GGFFT-RSAVQKF-- 762
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2432 QKAQAMISIVDIAT-----NQDENIATSSGNNFGDLKADKITYFGGLPTLRNLISMKARP 2486
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 762 -VQDPKYIGILLTHDNETYLQDDWYLISKVENSEPDIIFYUYKKRNDAMMGVGGSVLY 820

```

Db 2487 EVNLKRYSGCL-----KDIEISRTYPNLTSS-----PDYVG-----TKGCSLENNY 2528
QY 821 TRSAVLPESTIIPLOTAQGVGRDENTFIKTDNCG-----PEPLVE----- 864
Db 2529 TVS--FPKPGFVLSVPIDVGEINLSESTKNESGILLGSGGPAPRRRRROTQAY 2586
QY 864 -----RLEKKVEEGERTIIEVEEIEEVEKVRD-KEVTLFSKLEFGFELQDE-E 913
Db 2587 YVILLNGRLEVLHSTGARIMRKIV--IRPEPNLFHDGGRHSVHVERTGIFTVOVDENR 2644
QY 914 NPLRELKSEMDVLDGKMEATEVEKLF-GRALP 946
Db 2645 RYMONLIVE-----OPIEVKKLFVGAPP 2668

RESULT 10
US-08-125-077-2
Sequence 2, Application US/08125077
Patent No. 5872231
Patent No. 5872231 5840863
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
TITLE OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,077
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1130 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-125-077-2

Query Match 1.5%; Score 115.5; DB 2; Length 1130;
Best Local Similarity 20.0%; Pred. No. 0.46;
Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;
QY 294 NEFLHYQDDWYLLSSQIEKKRPDIYFYVYRGNDAMDGYGGSVITYTRSPILPE-SIIPN- 352

Db 6 NDVENEDHNLGKTRLENA-----DARNGDLLRLINPLGLSLAIPND 49
QY 352 ----LQAAKSVGDENFNFTITDNSCGPEPLVERLEKTAEEGKELIKEAVEIEEVEER 407
Db 50 TAAKLO-AVKDKAQAQAND--TAKDYLAQIITELHQNIDGLKKNKNKL--ADSAKTAAYAK 104
QY 408 EVER---VROTEML-----FORLEGEKELQODEENFRELSEKEKELINELQATE 458
Db 105 DPSKNKTIADADAVKKNLEQADRLIDKLPKIRELBDNKKNIS-EIKELINQARKOANS 163
QY 459 VE---KLFGRALPIKRLMALAPHSNLANET-----IKYVGS-----KLPH 500
Db 164 IKVSVSSGGGCIIRYKEIKKGSYNNIVVAVKTAVDNLFYGSKFIIDFLAIEKRKK 223
QY 501 KRESGW-----EDYFGSIYVAKICSSRIIPRYFRKSPRICGLDSRGLQ 545
Db 224 VSEFLMDVSGSVGRYEPDLTIDDSYWRIVASRTGRNGTI-----SVR---ALDGPAS 274
QY 546 LFSHGKHLSPASHINQNPVKGSGCKFPKDVALLMWEKMGQAKATAYAFILSYASA 605
Db 275 IVSTHSTSP-----PEYTLIDV-----DANMLFVGGLTGRL 308
QY 606 DAVDLKTCYCLKECRLELAKCI-----SNPACANVACLOTQNNRDECOITKGD 659
Db 309 KKADAVAVIT-----FTGCMGETYPDNR-----IGLWFRKEGDK 347
QY 660 LFENSVDDEFNECAVSRKKCVPRKSD-----VGDFPV--DPSVLYOKF-- 702
Db 347 -----GCTVS-----POVEDSEGITQFDGVALVSRPIRWYNISTVMEKFT 390
QY 702 -----DMKDF-----SGKW-----FI 712
Db 391 FSSALLMYLATDLDKREMSVELTDGHIKVSVDLGSGMASVSNQNHNDGKKSFTLSRI 450
QY 713 TRGLNPTFDAFDQOLHEFTEENKLV-----GNLSWRIRTPD-----GGFT-SSAVOKF-- 762
Db 451 OKANISIVIDF-----NOENIATSSSGNNRGDLKADKDIYRGULTNLNLSKKAP 505
QY 762 -YDDPKPGILYHNDDEYLLYQDDWYLLSSKYENSPEDIYFYVYRGNDAMDGYGGSVLY 820
Db 506 EVNLKRYSGCL-----KDIEISRTYPNLTSS-----PDYVG-----TKGCSLENNY 547
QY 821 TRSAVLPESTIIPLOTAQGVGRDENTFIKTDNCG-----PEPLVE----- 864
Db 548 TVS--FPKPGFVLSVPIDVGEINLSESTKNESGILLGSGGPAPRRRRROTQAY 605
QY 864 -----RLEKKVEEGERTIIEVEEIEEVEKVRD-KEVTLFSKLEFGFELQDE-E 913
Db 606 YVILLNGRLEVLHSTGARIMRKIV--IRPEPNLFHDGGRHSVHVERTGIFTVOVDENR 663
QY 914 NPLRELKSEMDVLDGKMEATEVEKLF-GRALP 946
Db 664 RYMONLIVE-----OPIEVKKLFVGAPP 687

RESULT 11
US-08-125-077-4
Sequence 4, Application US/08125077
Patent No. 5872231
Patent No. 5872231 5840863
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
TITLE OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA

ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125, 077
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3111 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-125-077-4

Query Match 1.5%; Score 115.5; DB 2; Length 3111;
Best Local Similarity 20.0%; Pred. No. 2.4;
Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;

QY 294 NEFLHYDDWYLLSSQIEKRPDDYIFVYRGNDAWDGYSVITRSPTLPE-SIIPN-352
DB 1987 NOVKNEDHNLKTRIEA-----DARNGDLTLMDTLGKLSAIRND 2030
QY 352 ---LQKAASVGRDFNNFITDSCGPPPLVERLEKTAEEGKLLIKAVEIEEVEK 407
DB 2031 TAKIQ-AVKDKARQAND--TAKDVLAQITTEHQLMDGLKKNYKRL--ADSVAKTNAVYK 2085
QY 408 EYEV---VVDTEML-----FORLLEGKELQDDEENFVRLSEKEELINELQWATE 458
DB 2086 DESKNKIADADATYKNLEQEARDLIDKLPKELEDNKKNIS-EIKELINQARKQANS 2144
QY 459 VE---KLEGRALPIRKLRLMALPHSNFLANHET---IKYVGS-----KLGH 500
DB 2145 IIVSVSSGDCIRTKPELKKSNNIYVNVKTAVADNLTFLGSAKFIIDLAITEMRGK 2204
QY 501 KPPSWG-----EDYFGSLVAKICSSRRIPRYEKKSPRICCGIDSRGLQ 545
DB 2205 VEFIMVSGVGRVPELDITIDSYWYRIVASRTGRNGI-----SVR--ALDGPAS 2255
QY 546 LSHGHNHNSPAHSINONPKNSGCKFPKDYALMWEMWGFATATVAITILSVASKA 605
DB 2256 IVPSTHSTSP-----PGYILIV-----DANAMELVGGLTGKL 2289
QY 606 DAVDAKTCTCLKRELELAKCI-----SNPACAANVACLOTGNNPDEDECOKGCD 659
DB 2290 KKADAVRVT-----FTGCMGETTYFDNKP-----IGLMNFKREKEDCK-- 2328
QY 660 LEENSVDPEFNCVSRKCVPRKSD-----VSGDFPV--PDPSVLVQKF-- 702
DB 2328 -----GCTVS-----POVEDESEGTIOFDGEGYALVSRPIRMYPNISTIVAFKFR 2371
QY 702 -----DMKDF-----SGKW-----FI 712
DB 2372 FSSSALLMTLATRDLRDFWSVLLTDGHIKVSYDLSGNASVSNONHNDGKKKSTLSRI 2431

QY 713 TRGINTPDAFDCQLHEFTEENKLV-----GNLSWRIRTPD---GGFT-PSAVQKF-- 762
DB 2432 QKQANSIVDIOT-----NEENIATSSSGNNNGDLKKADKKIYFGGLTLNLSKKAP 2486
QY 762 -VODPKYPGILYHNDYELLYODDWYLLSSKYENSPEYIFVYKGRANDAMPDGGSVLY 820
DB 2487 EVNLAKKYSGL---KDIEISRPYNILSS-----PDYGV-----TKGCSLENY 2528
QY 821 TRSNAVPESTIPLQTAQAVGVGRDFNTFIKEDTCG-----PEPPLVE----- 864
DB 2529 TVS--FPKPGFVELSPVPIDVGEINLSFTKNSSIIILGSGGTAPAPRRRRQGTQAY 2586
QY 864 -----RLEKKEVEGEERTIIEVEIEEVEKVRD-KEYTLFSKLFEGKELQDE-E 913
DB 2587 YVILLNNGRLFEVILSTGARIMRKIV--IRPEPLFHGDGSHVHVRGTGIFTVOVDENR 2644
QY 914 NLERLSKPEMDVLDGKMEATEVEKLF-GRALP 946
DB 2645 RYMONTLVE-----OPIEVKKLFVGAP 2668

RESULT 12
US-08-568-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121,001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-4

Query Match 1.5%; Score 115; DB 2; Length 1435;


```

QY 845 ENT-FIKDNTGCGPEPLVERLEKKEVEGERTIIKEVEIEEVEKVR-DKEVT---LF 898
DB 1206 YNAQIVOLE-----AMRNKELKLOESE-----KEKECLOHELOTTIGDLETSLQDMQ 1254
QY 899 SKLFEFKELOD-EENFL---RELKSEEMD-----VLGDKMEATEVEKLFGRALPIR 948
DB 1255 SOEISGLKCELDIAEKEKISGPHELSTSONDNAHLQCSLOTMTNKLNELEKIC-ELIOAE 1313
QY 949 KLMAYA-----THCFTS-----PCHDRIRFSSDDGIGRLGTRKRINGTFLTKI 993
DB 1314 KEVELTELNDSSRECIITARRKMAEEVGKLLINDDSGL-----LHGE-LVED 1363
QY 994 LPIPSADLRTTGG-----RSSRPLSAFRSGFSKIGDIYLPSPKNEKELTALLLKL 1047
DB 1364 LP-----GGEFGEDPNEOHVYS-----LAPLDSNSYEHLT---LSOKE 1399
QY 1048 VGV---LACAPLIVPS-----ADAVDALKTACLLKGCRIELAKCIAN--- 1088
DB 1400 YOMHFAELQEKFLSLQSEHKILHDQCMSSKMSLOTYVDSLKAENLYLSTLNRFQGD 1459
QY 1088 -----PACANAACTCNCNNRPDETCQIKGDL-----FENSVYDE 1124
DB 1460 LYKEMOLGLEEGLVPSLSSSCVPSDSSLSLSDSSFYRALLBQTDMSLNLNLEGA---S 1517
QY 1125 FNECAVSRKKCVPRKSDLGEPAPDPVLYVONFNISDENG-----KMYITSGLNPTFA 1178
DB 1518 ANOCSVDEVCSSLOTIVYSLAEN---LVLSTNLNRFQGDLYKEMOLGLEEGLVPSLSS 1574
QY 1179 F-----DCOLHEFTE--GDNKLTVGNISWRK---TLDGSEFT----- 1212
DB 1575 SCVPDSSLSLSDSSFYRALLBQTDMSLNLNLEGA---S 1634
QY 1212 -----RSVQKFYQDPNRCVLYNHNENIHYODWYILSKIE 1250
DB 1635 KETPSAPARAVEELSELCVYRQSEKLEKMEKESQIMNKKEIOLE---QLSS--E 1687
QY 1251 NKPEDIIFYYRGNDAMWGYGAVYTRSSYL-----PNSIIPLEKAASI-GRDPS 1303
DB 1688 RQILDRLKQOYLSNENOCOKLTSTLEMEBSKLAARKOTBOLSTLEVARLOQLDLS 1747
QY 1304 TELRDNTOGPEPALVERIEK-----TVEGERIIVKEVEI-----EEVEKEVEKY 1351
DB 1748 S--RSLGIDTDEAIGRNECISOISKEHSETERTPKHDVHQCCKDAQOOLNIDIEKI 1805
QY 1352 GRT-----EMTLQRLAEGFNEI-----KQ 1371
DB 1806 TETGAVKPTGECGSEOSPDNTYEPGEDKTI--OGSECEISLSSFGPNALVPMDELQNOE 1863
QY 1372 DEENF---VRELSEKMEFLDEIKMEASEYEKLFGR 1404
DB 1864 DHHNOLRYKETSNNRLHLVIEDRDRKVESILNE 1899

```

RESULT 14
PCT-US95-16216-1
Sequence 1 Application PC/TUS9516216

GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
APPLICANT: Rattner, Jerome B.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
NUMBER OF SPOUNCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
PCT-US95-16216-1

```

Query Match 1.5%; Score 112.5; DB 3; Length 3248;

Best Local Similarity 17.7%; Pred. No. 4.6; Indels 431; Gaps 60;

Matches 219; Conservative 178; Mismatches 408;

QY 381 ERLKTAEE--GKRLIKAEVEIEEVEKEVEKYRD---TEMTLQRLT-----EGFK 428

DB 883 QRIKLEDDTSARHNVAVAFLLSLENKEKELQLLNDKVEFQAEIQELKSNLLDLSK 942

QY 429 ELQDEENFVRELSEKEKEILNLOMEATEVEKLFGRALPIRLKRLMALPHS--NFLANHE 487

DB 943 ELQLLSET---LSLEKREKSSIIISLNKREIEELTOENGLIKEINASLNQEKNNLIQKSE 998

QY 488 TIKYVGSKLPGHRRFSWGEDYFGSIVAKICSSRRIPRYFFKSPRICCGDLSRGLQLF 547

DB 999 SFANYI----- 1008

QY 548 SHGHNLSPASHINQNPKNCGCKPPKDYALMWKMGQFAKTAIYALFISVASKADA 607

DB 1008 -----EKSISELDQYKQKLLILQRCETGN-----AYEDLSQKXKA-A 1046

QY 608 VDALKTCCTCLLKECRLELANCISNPACANAVACLOTNNRPDETCQIKGDLFENSVD 667

DB 1047 QEKNSKLECLINE-----CTSLCENKNELE-OLK--EAFKAEHOE 1084

QY 668 FNECAVSRKKCVPRKSDVGDVPYDPDSVLYOKTDMKDFSGKVFITRGLNPTDADCOL 727

DB 1085 FLTKLAFAEER-----NONLME-----LETVOQLRSEM 1114

QY 728 HEFTENKLVNLSWRIRPDGFFTRSAOKFVOPDKYGLIYNHNDNYLYODWYI 787

DB 1115 TDNONNSKSEAGLQKELMTLKE---EONKQKQEVND-----LLOENQOLM---KV 1159

QY 788 LSKVE--NSPEDIIFYYRGNDAMWGYGGSVLYTRSAVLPESIIPELOTAQAKYGRD- 845

DB 1160 MKTKHECQNLSEPTI-----RN-----SVKERESENNOCNFKPQMDLEVEKISLDS 1205

QY 845 ENT-FIKDNTGCGPEPLVERLEKKEVEGERTIIKEVEIEEVEKVR-DKEVT---LF 898

DB 1206 YNAQIVOLE-----AMRNKELKLOESE-----KEKECLOHELOTTIGDLETSLQDMQ 1254

QY 899 SKLFEFKELOD-EENFL---RELKSEEMD-----VLGDKMEATEVEKLFGRALPIR 948

DB 1255 SOEISGLKCELDIAEKEKISGPHELSTSONDNAHLQCSLOTMTNKLNELEKIC-ELIOAE 1313

QY 949 KLMAYA-----THCFTS-----PCHDRIRFSSDDGIGRLGTRKRINGTFLTKI 993

```

Db 1314 KYELVTEINDSRSECTATRKMAEYKLLINEKLLINDSSGL-----LHGE-LVED 1363
QY 994 LPTIOSADLTRTGG-----RSSRPLSARSGSGKIGDIVPLPSKNELKELTAPLLKL 1047
Db 1364 IP-----GGEGEOPNEOHVPS-----LAPLDESNSYEHLT--LSDKE 1399
QY 1048 VGV-----LACAPLIVPS-----ADAVDALKTGACLLKGRILEACIAN----- 1088
Db 1400 VQMHFALEDEKFLSLQSEKILHDQCQSSKMSLQITVDSLKAKENLVLSTNLNFGOD 1459
QY 1088 -----PACANVACLOTCNNRPDETCQIKCGDL-----FENSVDYE 1124
Db 1460 LVKEMQLGLEGLVPSLSSGCVDDSSLSLGGSSFYRALLEQTDMSLSTLBEAV--S 1517
QY 1125 FNMCASRRKKCVPRKSDJSEFPAPDSVLYQNFNLSDENG-----KWITISGLNPTDA 1178
Db 1518 ANQCSVDEVFCSSLQTYVDSLAKEN--LVISTNLNFGODLVKEMQLGLEGLVPSLSS 1574
QY 1179 F-----DCQLHFEHTE--GDNKLVGNISWRK-----LTDGHEFT----- 1212
Db 1575 SCVPSDSSLSLGGSSFYRALLEQTDMSLSTLBEAV--S 1634
QY 1212 -----RSAYQEVQDPNQPQVLYNHDNEXLYHQDDWYLLSKIE 1250
Db 1635 KETPSAPAKGVELESLCEVYRQSLKLEKKMESOGIMKNKEIPELE-----QLLSS--E 1687
QY 1251 NKPEDIIFYRRGNDNMGYGGAVYVTRSSYL-----PNSITPELEKAKAST-GRFSS 1303
Db 1688 ROELDLCKROYLSENQOWOQKLTSLVLEMSKLAERKOTEOLELHVARLOLOGDLS 1747
QY 1304 TETFRDNTGPEPALVERLEK-----TVEGERIIVKEVEEI-----EEVEKEVEKY 1351
Db 1748 S--RSLIGIDTDAIQGRNESCDISKHESTETTERPKHDVQIODKQAQDNLNDIKI 1805
QY 1352 GGT-----EXTLFORAEGFNEI-----KQ 1371
Db 1806 TETGAVKPTGEGSGEOPDPTNYPEGDEDT--QGSSECSLSLSSFGPNALVPMDFLGNQE 1863
QY 1372 DEENF--VRLSKEMEFLDEIKMEASEVEKLFK 1404
Db 1864 DIHNLOLRKETSINENRILHVIEDRKRKVESLNE 1899

```

RESULT 15
US-08-603-753D-4

Sequence 4, Application US/08603753D
Patent No. 5891857

GENERAL INFORMATION:

APPLICANT: HOLT, JEFFREY T.
APPLICANT: JENSEN, ROY A.
APPLICANT: KING, DAVID L.
APPLICANT: KING, MARY-CLAIRE
APPLICANT: SZABO, CSILLA I.
APPLICANT: JETTON, THOMAS L.
APPLICANT: ROBINSON-BENTON, CHERYL L.
TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARLES A. TAYLOR, JR.
STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
CITY: DURHAM
STATE: NORTH CAROLINA
COUNTRY: USA
ZIP: 27707

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 800 KB storage
COMPUTER: IBM PC/XT/AT compatible
OPERATING SYSTEM: Windows 3.1
SOFTWARE: WORD PERFECT 6.1 and ASCII

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603,753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: protein
TOPOLGY: unknown
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-08-603-753D-4

Query Match      1.5%  Score 111.5; DB 2; Length 3418;
Best Local Similarity 17.8%; Pred. No. 6.1;
Matches 275; Conservative 212; Mismatches 576; Indels 483; Gaps 72;

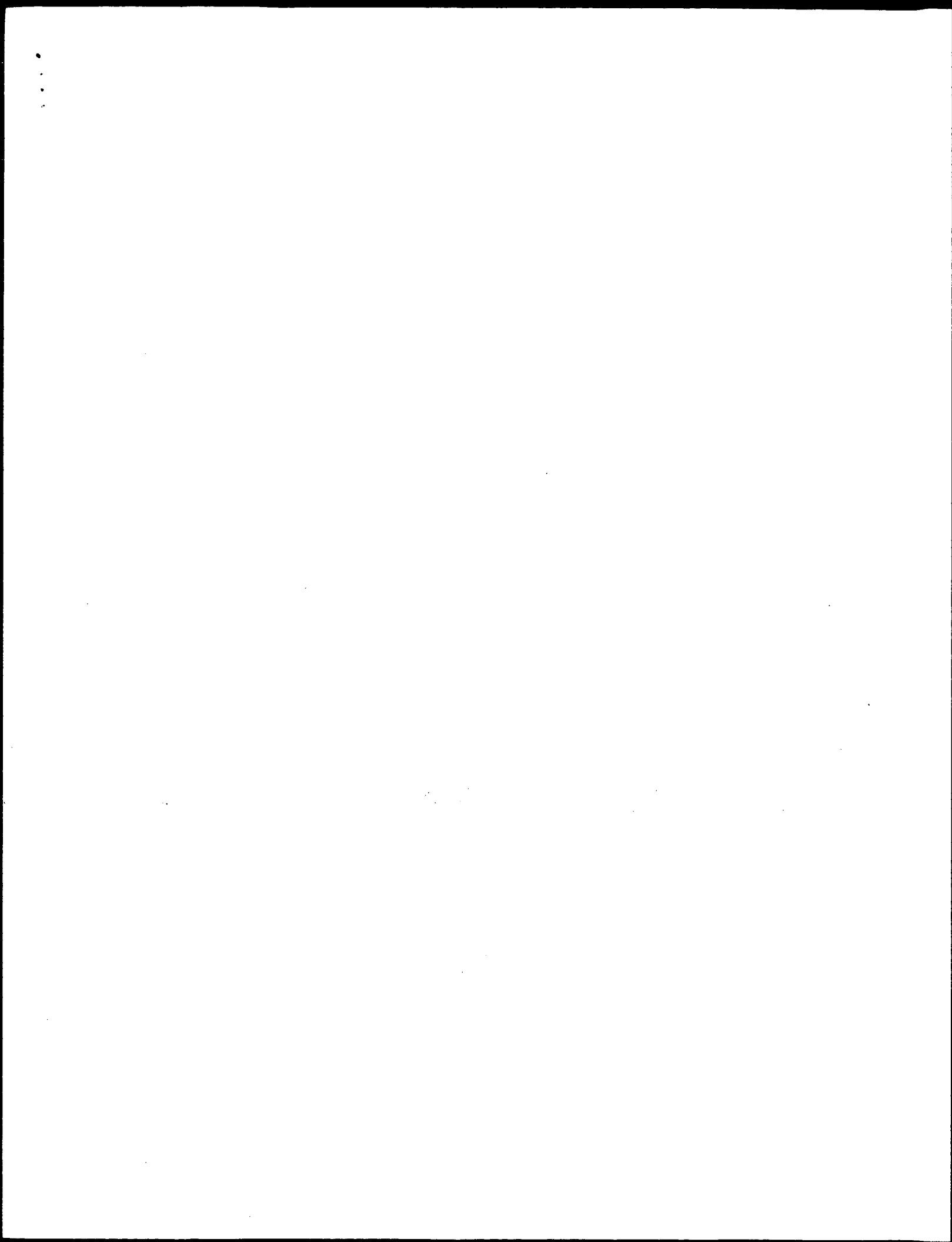
QY 3 LSHYTFLEKKEALNLYARSPCNERFHRSGOPT---NIMMKIRSNNGYNSRPLFTSY 59
Db 540 LEHYVCSQKEDSL-----CPNIDNGSWPATTTQSVALK--NAGLSTLTKKTKNK 589
QY 60 KTSFSDSHCK-----DKSQICSDTSEFEIQRFLKGMILLLEKQWRQFOILAIV 112
Db 590 FYAIHDEFFYGKKIPKQKSELINCSAQF-ENAFE---APLTFANADSGLLHSYVK 644
QY 113 LVCTIYIVRVAVNALKTCACLLKECRILEIAKCIANDSCAANVACLOTCNNRPDETECO 172
Db 645 RSCS-----QNDSEPTLSLTSSFGITILKCSRNTECSNNTVISODLDYK--EAKCN 694
QY 173 IKCGDLFENSVDYD-----QFNCAVSRKKCVPRKSDVGEFVY-----PDRNAVQNMNMK 222
Db 695 KRKLQFLFPEADSLSCQBOGCENDPKS--KYSVDIEEVLAAACHPVQHSKVE--ISDT 751
QY 223 DFGSKWYI-----TSGLNPTFDAPDCOL-----HEFMENDKLVGNLTWRIKTLDG 268

```

Db 752 DFOSOKSLIYHENAUSTILPTSKDVLSNLVMSRGESEYKMSDKLGN----- 802
QY 269 GEFTRSAVOIFVODPDLPALYNHNEFLAYODDWMILSSQIEN-----KPDYIFVYRG 324
Db 802 -----ANESVDLKNIPMEKNQ-----DVICALNENYKVELLPPEKYRVAVSPS 846
QY 325 RNDAMDYGGSVYITRSTPESTIIPNLOKAKSVGR-----DFNNFI----- 368
Db 847 R-----KVOFNQNTNL--RVIOKNOEETTSISKITVNPDESEELFSNNENFVQVA 895
QY 368 -----ITDNGSGPEPLV-ERLEKTAEEGR-----LLINEAVEIEEVE 406
Db 896 NERNNALGNTKEIHEITDTCVNEPIFNKSTMVLYGDTQKQATQVSIKQDLYVLAEN 955
QY 407 KEVEKRDTEMTLFOPL-----LEGFRELQDEEN-----FVRELK 443
Db 956 KNSYK-QHIXKMTLQODLKSIDLNDKPEKNNDYMNKMACLLGPISNHSGSGFRASN 1014
QY 444 EKEEI---LNELOMEATEVEKLEGRALPIRLKLMALA-PHSNFLANHETIKYVGSKL 498
Db 1015 KEIKLSEHNKSKMFEKDIEQYPTSLACVEIYNTLALDNOCKLSKPSI-----NTVS 1069
QY 499 GHKRFSGWGEYFGSIYVAKICSSRIPIRKRSPRICCGILDSRGLOFS---HGKHL 554
Db 1070 AHLQ-----SSVYVSDCKNSHITPO-----MLFSKODFNSNHL 1103
QY 555 SP---AHSIGNONVPRKNGCGKFPDVALMWEKMGOKATAIVAFILSVAS---KADAV 608
Db 1104 TPSQAKETTELSTIEBGSOF-----EFTQFRKPS---YILOKSTFEVPEKNQ 1149
QY 609 DALTCITCLKECRLELAKCISNPACANVACLOTCNNRPDETECOIKCGDLFENS--- 665
Db 1150 TILKTS---BECDRADLHVIMN---AFSIGVDSKOFEGTVEIKRRFAGLLKNDCKNS 1203
QY 665 ----VUDE-----FNECAVSRKKCVPRKSDYGDVFPVDPVSVLYOKFMDK 705
Db 1204 ASGYLTDENEVGFPGFYSANHTKLNSTALQAKVCLFSDIENIS-BETSAVHPISLSS 1262
QY 706 FSGKWFITRGINPTFADFCOLHEFT---DENKLVGNLSWRIRTPDGGFTRSAVOKFV 762
Db 1263 -----SKCHDSYVSMKTIENHNDKTYSEKNNKCOLIIONNITEMTIGTF-----VEEIT 1310
QY 763 QDPKYPGILYNHNEVLLYODDWMILSSKVENSPEDYIFVYKGRNDAMDYGGSVLYTR 822
Db 1311 EN--YKRNTEENEDNKYTAASRNSHLEFDGSDSKNDYVCJHKDET-----LLFTD 1360
QY 823 SAVLPESIIPLOTRAAQVGRDNTFITD-----NTCGPEPPLVER 864
Db 1361 Q-----HNICLKLSGQFMKEG---NTQIKEDLSLFLLEVAKAQACHGNTSNKEQLTATK 1413
QY 865 LEKRYEGERTIIEVEEIEEVEKVRDKEVT---LFSKLEGEFKELORDEENFL--- 917
Db 1414 TEQN-----IKDFETSDIFFOQASGNISVAKELFNKIVNFFDQKPELHNFSLNSE 1465
QY 917 --RELKSEEMVDLQKMEATEV--EKLFGALPPIRLKLMATAVTHQFTSPCHDIRFFSSD 972
Db 1466 LHSDIRKMKMDI---LSEETDIIYKHKILKESVY----- 1498
QY 973 DGIGRLGIT-----KRINGTFL-----KILPPIOSAD---LRTTGGRSSRPL 1013
Db 1498 -GTGNOLVTFQOGQPERDEKIEPTLLGFHTASGKVKJAKESLDKVKXNLFDEKEQGTSEI 1556
QY 1014 SAFRGSGFGIFDIYLPKSNELKELTAPLLKLGVYLACAFLLVPSADAVDAKLTQACL 1073
Db 1557 TSFSHQWAK-----TLKYREACDLE-----LACETIETTAAPKCKEMQN--SL 1598
QY 1074 LKGCRIELAKCIANPACANVACLOTCN-----NRPDET-ECQJIC--- 1114
Db 1599 NNDKMLVSIETVYPPKLLSDNLCROTENLKTSTIFLKVVENVERKETAKSPATCYTNQ 1658
QY 1114 --GDLFENSVDVEFNECAVSRKKCVPRKSDJGEPAPDPVSVLYQNFNSDNGKWIYITSG 1171
Db 1659 SPYVSENSALAFYTSQ--SRKTSVSTSL-----EAKRWMLREG 1696

QY 1172 LNPTFADPCOLHEFTEDGNKLVGNISMRITKTLDSGEFTRSAVOKFVODPNQGVLYNH 1231
Db 1697 I-----FDGQPE-----RINTAD-----YVGNLYENNSNSTIEN 1727
QY 1232 DNEYLHYQDDWMILSSKLEKNKPEDYIFVYRGRNDAMDYGGAVVYTRSSVLPNSIPEL 1291
Db 1728 DKNHLSKQDITLSSNSMSNS-----YSYHSDVYNDSG---YLSKNKLDGIEPVL 1776
QY 1292 EKAASIGRDFSTFIKTDNTCGPEPALVERIKETVEEGERIIVKEY 1337
Db 1777 KNEVDOKNTSFSKVISNKKDANAHYPOTV-----NEDICVEEL 1813

Search completed: October 14, 1999, 00:47:04
Job time: 755 sec



Page 1

```

1
US-08-232-463-14/c
: Sequence 14, Application US/08222463
: Patent No. 5670367
:
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: City: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patatin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232.463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935.313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
:
: INFORMATION FOR SEQ. ID NO.: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: lin. nar
: IMMEDIATE SOURCE:
: CLONE: PTZ9pt.F15
:
: US-08-232-463-14

```

ALIGNMENTS

Sequence 8, Appli
Sequence 21, Appli
Sequence 21, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 10, Appli
Sequence 1, Appli
Sequence 4, Appli

[illegible]

```

1      RESULT 3
2      US-08-676-974-2
3      : Sequence 2, Application US/08676974
4      Patent No. 5770422
5      GENERAL INFORMATION:
6      APPLICANT: COLLINS, KATHLEEN
7      TITLE OF INVENTION: Human Telomerase
8      NUMBER OF SEQUENCES: 10
9      CORRESPONDENCE ADDRESS:
10     ADDRESSEE: Science & Technology Law Group
11     STREET: 268 Bush Street, Suite 3200
12     CITY: San Francisco
13     STATE: CA
14     COUNTRY: USA
15     ZIP: 94104
16     COMPUTER READABLE FORM:
17     MEDIUM TYPE: Floppy disk
18     COMPUTER: IBM PC compatible
19     OPERATING SYSTEM: PC-DOS/MS-DOS
20     SOFTWARE: PatentIn Release #1.0, Version #1.30
21     CURRENT APPLICATION DATA:
22     APPLICATION NUMBER: US/08/676,974
23     FILING DATE:
24     CLASSIFICATION: 530
25     ATTORNEY/AGENT INFORMATION:
26     NAME: Osman Ph.D., Richard A
27     REGISTRATION NUMBER: 36,627
28     REFERENCE/DOCKET NUMBER: DCB96-055
29     TELECOMMUNICATION INFORMATION:
30     TELEPHONE: (415)343-4341
31     TELEFAX: (415)343-4342
32     INFORMATION FOR SEQ ID NO: 2:
33     SEQUENCE CHARACTERISTICS:
34     LENGTH: 2277 base pairs
35     TYPE: nucleic acid
36     STRANDEDNESS: double
37     TOPOLOGY: linear
38     MOLECULE TYPE: CDNA
39     :
40     :
41     :
42     :
43     :
44     :
45     :
46     :
47     :
48     :
49     :
50     :
51     :
52     :
53     :
54     :
55     :
56     :
57     :
58     :
59     :
60     :
61     :
62     :
63     :
64     :
65     :
66     :
67     :
68     :
69     :
70     :
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :

```


OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 2.8%; Score 43.2; DB 3; Length 32207;
Best Local Similarity 47.7%; Pred. No. 0.012;
Matches 126; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1198 aagaaggttgaagagatagaagaaggttagaagaaggttgaagaaggttcgtagaagctg 1257
DB 20903 ACAGAGGGGATGACGAGAGAGATGACGAGAGATGACGAGAGAGATGACGAGAGAGATG 20844
QY 1258 agatgaactgttccagaagattgctgaagatttaactgaagcagaagcagaagaga 1317
DB 20843 ACAGAGGAGGAGGACGAG 20784
QY 1318 attcgtgaagaaggttaagtaagaagaagattgatttggatgaagtcataatggaag 1377
DB 20783 ATGACGATGATGAGGACAAATGAGAGAGAGAGATGACGAGAGAGAGAGAGAGAGAG 20724
QY 1378 caagtgaagttgaaataattgtttggaagaagcttgcgaatcagaaggtcgaagtgaac 1437
DB 20723 ACAGAGAGGAGCGGGGCGATGGAACAAACCTTGAGCATCCAAAGTTCAACAAGCAGC 20664
QY 1438 aagaaccaccattgttgtacaac 1461
DB 20663 AGGAGCCACACAGCAGAGCCAC 20640

RESULT 6
US-08-728-323A-1

Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/NSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 2.8%; Score 43.2; DB 4; Length 3489;
Best Local Similarity 47.7%; Pred. No. 0.0033;
Matches 126; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1198 aagaaggttgaagagatagaagaaggttagaagaaggttgaagaaggttcgtagaagctg 1257
DB 1094 ACAGAGGAGATGACGAGAGAGATGACGAGAGATGACGAGAGAGATGACGAGAGAGATG 1153
QY 1258 agatgaactgttccagaagattgctgaagatttaactgaagcagaagcagaagaga 1317
DB 1154 ACAGAGGAGGAGGACGAG 1213
QY 1318 attcgtgaagaaggttaagtaagaagaagattgatttggatgaagtcataatggaag 1377
DB 1214 ATGACGATGATGAGGACAAATGAGAGAGAGAGATGACGAGAGAGAGAGAGAGAGAG 1273
QY 1378 caagtgaagttgaaataattgtttggaagaagcttgcgaatcagaaggtcgaagtgaac 1437
DB 1274 ACAGAGAGGAGCGGGGCGATGGAACAAACCTTGAGCATCCAAAGTTCAACAAGCAGC 1333
QY 1438 aagaaccaccattgttgtacaac 1461
DB 1334 AGGAGCCACACAGCAGAGCCAC 1357

RESULT 7
US-08-323-170B-1

Sequence 1, Application US/08323170B
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williams, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf230
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

Query Match 2.6%; Score 40.2; DB 2; Length 9636;
Best Local Similarity 49.8%; Pred. No. 0.046;
Matches 102; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 1173 gaagaagtgaagaagataatcgtaaaagaggttgaagaagatagaagaagagtagaag 1232
DB 1298 GAGGTAGAGTGAAGAGGTAGGTGTAAGAGGTGTAAGAGGTGTAAGAGGTGTAAG 1357
QY 1233 gaagtggaaagagtcgttagagctgaatgacctgttccagaagatttgctgaagattt 1292
DB 1358 GAGGTAGAGTGAAGAGGTAGGTGTAAGAGGTGTAAGAGGTGTAAGAGGTGTAAG 1417
QY 1293 aatgactgaagaagaagagagaggaatttcgtgaagaagattgaagaagaagagtagag 1352
DB 1418 GGGGTAGGTGAAGAGGTAGGTGTAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGA 1477
QY 1353 ttcttgatgagatcaaatggaag 1377
DB 1478 TATGTAGATGAAGAAAGCAAGCAAG 1502

RESULT 8
US-07-865-662F-10/C
Sequence 10, Application US/07865662F
Patent No. 5451670
GENERAL INFORMATION:
APPLICANT: Marcia M. Miller
TITLE OF INVENTION: Restriction Fragment Length
POLYMERIZATION Test For Haplotyping Domesticated Fowl
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: City of Hope
STREET: 1500 East Duarte Road
CITY: Duarte
STATE: California
COUNTRY: United States of America
ZIP: 91010-0269
COMPUTER READABLE FORM:
MEDIUM TYPE: 3M Double Density 5 1/4" diskette
OPERATING SYSTEM: MS DOS Version 3.20
SOFTWARE: Microsoft
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/865,662F
FILING DATE: 07 April, 1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/688,326
FILING DATE: 22 April 1991
APPLICATION NUMBER: 07/588,922
FILING DATE: 27 September 1990
APPLICATION NUMBER: 07/210,405
FILING DATE: 23 June 1988
APPLICATION NUMBER: US 07/130,529
FILING DATE: 5 December 1987
APPLICATION NUMBER: US 07/068,176
FILING DATE: 30 June 1987
ATTORNEY/AGENT INFORMATION:
NAME: Icons, Edward S.
REGISTRATION NUMBER: 16,541
REFERENCE/DOCKET NUMBER: No. 5451670e
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 783-6040
TELEFAX: (202) 783-6031
TELEX: No. 5451670e
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE: Synthetically Prepared
IMMEDIATE SOURCE: Synthetically Prepared
US-07-865-662F-10

Query Match 2.5%; Score 39; DB 1; Length 2188;
Best Local Similarity 61.2%; Pred. No. 0.045;
Matches 63; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1148 cgtggagagaattgagaagacagtggaagaagtggaagaagataatcgttaagaaggttga 1207
DB 184 CGAAGAGAGCAAG 125
QY 1208 agagatagaagaagagtagaagaagagtggaagaagtcggt 1250
DB 124 AGAGAGCAAGAT 82

RESULT 9
US-08-332-766A-22
Sequence 22, Application US/08332766A
Patent No. 5843647
GENERAL INFORMATION:
APPLICANT: JEFFREYS, Alec J.
TITLE OF INVENTION: SIMPLE TANDEM REPEATS
NUMBER OF SEQUENCES: 125
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3936
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,766A
FILING DATE: 01-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9326052.9
FILING DATE: 21-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: BIRD, Donald J.

1 PRIOR APPLICATION DATA:
2 APPLICATION NUMBER: US 08/123,933
3 FILING DATE: 17-SEP-1993
4 CLASSIFICATION:
5 ATTORNEY/AGENT INFORMATION:
6 NAME: LAZAR, Steven R
7 REGISTRATION NUMBER: 32,618
8 REFERENCE/DOCKET NUMBER: 5203-PCT
9 TELECOMMUNICATION INFORMATION:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525


```

US-08-474-503-1
: Sequence 1, Application US/084/4503
: Patent No. 574446
: GENERAL INFORMATION:
: APPLICANT: Emory University
: TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kilpatrick & Cody
: STREET: 1100 Peachtree Street, Suite 2800
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: US
: ZIP: 30309
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,503
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Pratt, John S.
: REGISTRATION NUMBER: 29,476
: REFERENCE/DOCKET NUMBER: EM0106CIP(3)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-815-6500
: TELEFAX: 404-815-6555
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9009 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapien
: TISSUE TYPE: Liver
: FEATURE:
: NAME/KEY: misc_feature (Domain Structure)
: LOCATION: 5125 . . . 7053
: OTHER INFORMATION: /note= "Equivalent to the A3-C1-C2
: OTHER INFORMATION: domain"
: FEATURE:
: NAME/KEY: misc_feature (Domain Structure)
: LOCATION: 1 . . . 2277
: OTHER INFORMATION: /note= "Equivalent to the A1-A2 domain."
: FEATURE:
: NAME/KEY: Domain
: LOCATION: 1..2277
: OTHER INFORMATION: /note= "CDNA encoding human factor
: OTHER INFORMATION: VIII."
US-08-474-503-1

```

	Query Match	2.3%	Score 35.8	Dt 2,	Length 9009;
	Best Local Similarity	46.0%;	Pred. No. 0.91;		
Matches	121:	Conservative	0;	Mismatches 142;	Indels 0; Gaps 0;
OY	1292 taatgaactgaaagcaagaacgaagagaatttcgtgagagttaagtaagaagagatgta	1351			
Dd	7984 TAAAAAATAATCATCGAGGAGTCACATTCTTAAAGCCTTGAAAATTAATAATCACTGTCTT	8043			
OY	1352 gtttttgatggatatcaaatgyaaagcaagttagytltgaaaaattgtytggysaaagcttt	1411			
Dd	8044 CTTGAATTTTGTGATGGCCACAAGAAGAAATGATGTGCATTAGGCTCTTAAGSACT	8103			
OY	1412 gccaatccaggaaaggtcaggtlagaanaaacaccacctattgttgtacaacatactattatbc	1471			

Accession	Sequence	Position
D6	ACATTATATTTCTGTGGAATATGAGGAAATCCATGTTATCTGATGAGTGAATAC	8163
QY	1472 atactggttcggttcataaaagtaataattttgtacacagtcacatcatccataac	1531
D6	8164 AAATCTTGATATCTCTATATGACACACCAAGTTACTCTCTCCCTCTACATTAATTTCTGCTG	8223
QY	1532 aattgataaaaaaataaaaaa	1554
D6	8224 AAAATATACACACAAAATGTAA	8246

Search completed: October 14, 1999, 00:43:45
Job time: 3481 sec

•
•
•
•


```

Db 490 KSEAGGLKOEIMTLKEEQNMKEVNDLLOENQOLMKVKTKEQNESEPIR----- 544
QY 552 HNSPASHINONPKNSGCKFPKQVALMWKMGOFATVAIVAFILTSVASKADAVDAL 611
Db 544 -----NSVKERSEBNO-CNRPOMDLVEKSLSDSYNAOLVOLEAMRNKELKLOESE 596
QY 612 KTCTCL-----LKECRLEL-AKCTISNP-----ACAN 637
Db 597 KEKELQHLQIIRIGDLETSLNLODMQSOEISGKDEIDAEKELYSGPHELSTSONDAH 656
QY 638 VAC-LOTGNRPDETE--COIKGDLFENSVDNEFNECAVSRRKCV--PRK--SDVGF- 690
Db 657 LQCSLOTNNKINLEKIEIILQAEKYE--LVTEIEND--SRSECTATARKMAEVEGKLL 711
QY 690 ----PVEDSVLYOKFMDGSKWFTITRGILNP-----TDAADC 725
Db 712 NEVKILNDOSGLHGLVEDIDPGEF--GEOPNEOHPSVLAFLDESNSEHLTSDREV 768
QY 726 QLHEFTIEENKLVGNLSMRIRTPDGGFTPSAVOKFVODPKYPGILYNHNEYLLOD-- 784
Db 769 QMHFAELQK-----FLSLQSEHKLHGHQO--MSKMSLQTYVDSL 810
QY 784 --DWYLLSKVENSPEPDYFYVYKGN-----AMDYGGSVLYTRSAVL 826
Db 811 KAEIVLSTNLRNFOGDLVKEMQLGLEGLVPSLSSCVPDSSISLGDSSFY--RALL 868
QY 827 PE-----SITPELOTAO-----KVRDENTFKTDMTCPEPL-----VERLEKVEGE 873
Db 869 EOTGMSLLSNLEGAVSANOCSVDVFCSDLOENITREKTPSAKAKVEELSECEYR 928
QY 874 RTIIEVEIEEVEEK--VRQEVTLFSKLEP-----FKELQDEENFIELSK 922
Db 929 OSU-----EKLEKMEQOIMKNKEIQLEQLLSEEROLDCKRQIYLSNEQOQKLSV 984
QY 923 EMDVDGLMKEATVEYKLE-----GALPIRKIMANATP-----CFTS-- 961
Db 985 TLEMESKLAEEKQTEQSLLEVARLOGLDLSRSLGIDTDAIOGRNESDISKE 1044
QY 961 -----PCHDRIRFFSSDDIGRLGITRRKINGTEFLKLPPI-----QSADE-- 1003
Db 1045 HTSETTERPKHD-VHOLCKDAQODLNDIEKITETGALK--PTGECSSGQSDPTWYE 1100
QY 1003 -----RTTGGRS-----SRPLSAFRSGSKGIFDIYPLPSK----- 1034
Db 1101 PGEEDKTQSSSEICISELSEFSGNALVPMDFLGNQEDIHNLQRYKETSNNMLRLHYED 1160
QY 1034 -----NEIKELTAPRLKLVGVLAACAPLIVPSADAVALKACALLKGCRIELAKC 1084
Db 1161 RDRKVESLINKEKELDSKHLDOEVLMT-----KTEAC-----TELEKI 1199
QY 1085 IANPACAAVACLQTCNNRPDETECOIKGDLFENSVDNEFNECAVSRRKCVPRKSDG- 1144
Db 1200 V-----GEL-----KK-----ENSDISE 1212
QY 1144 --EPFAPDPVSIVQNPNTSDENGKWTYITSGNLPTDAPDCOLHEHTE--GONKILVGN 1197
Db 1213 KLEYFSCHQOELQREVTESE-----GLNS-----DLEMHADKSSREDIGNVAKVN 1258
QY 1198 ISMRITKIDSGFFTSRQVOKFVODNPQVLYNHNEYLH-----YQDDWYIILSKIEENK 1292
Db 1259 DSMKRFELD-----VENELSRIRSEKASIEHAIPLLEADLEVYQVEKCLEKDNENK 1310
QY 1253 P-----EDYIFVYVYRGRNDAMDYGAVVYTRSSVLPNSIIPLEKAKAKSIGRDFSTFI 1306
Db 1311 QKVIACLEBELSVTSENOL--RGELDTMSKKTALTADOLSEKKNKEKTOLESBHQSCCL 1367
QY 1307 KTDNCGEPALVERIEKTEVEGERIYKEVEEIEEF--VEKEVEYKGTENTLFLORLA 1363
Db 1368 HCIQVAAEVEKETELOTLSSVSELLKDKTHLOEKLOSLEKSOALSTLRCELENDIA 1427
QY 1364 EGFNLLKODEENFVE-----LSKEEMEFIDEIK--MEASEVVK 1400
Db 1428 ----QINKEKELLYKESSELOARLSESDYEKLNVSKALEALVER 1468

```

```

RESULT
2
US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Delinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50, pro-translation of seq ID NO:54
US-08-592-126-148

Query Match 1.8%; Score 135.5; DB 2; Length 1312;
Best Local Similarity 19.7%; Pred. No. 0.012;
Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

QY 382 RLEKTAEEB-----KLLIIEAVEIEEVEKEKVRDIENTLIFORLLGFELODEENF 437
Db 398 KLVREOBGBAKTANOLMNDFAKEKFLKOKOIDEIRD-KKTGLGRITELKSEI----- 450
QY 438 VRELSEKEEILNELOMEATEVEKLEFGALPIRKIRMLAPHSNPL-----ANHETIKY 491
Db 450 ---LSKKQ---NELKNVYELQOLEGSSDRILLEDQELIKARELSAENKSNVETIKM 502
QY 492 YGSKLPGHKRRFSWGWEDYFGSIIVAKICSSRRIPRYRKSFRICCGDLSRGLQIFSHGK 551
Db 503 EYIS-LQNEK-----ADLRITLRK-----LDQEWELQ-NH-- 531
QY 552 HNSPASHINONPKNSGCKFPKQVALMWKMGOFATVAIVAFILTSVASKADAVDAL 611
Db 531 -----HTTRQOMEKLTIDKADKDEQIKIKSRHSDELTSILGYF-----PNKQLEDM 580
QY 612 KTCTCLKECRLELAKCTISNPACAAVACLQTCNNRPDETECOIK-----CGDLFE 662
Db 581 HSKSKELINQTRIRAKL--NRELASSEQNNKINNELKRREDEQLSSYEDKLFVCGSDGF 638
QY 663 NSVDEFNECAVSRRKCVPRKSDVGDVPVDPDSVLYOKFDMKDSGKWFITRGILNPTFDA 722
Db 639 ESDLRLEK-----ETIEKSKQRAMAGATAVYSQ 668

```

Thu Oct 14 07:57:31 1999

us-09-075-375-6.rail

Page 3

```

QY 723 FDCOLHEHEENTLVNLSWIRITPDGGEFIRSAVOKEVODPEKPEGLYHNHNEYLLYO 782
Db 669 FIILO-----IDENSCCPVOCRV-----FQTEHLEGEVSD-----701
QY 783 DDMWTLSSKYENSPEDYIFV--YKGRNDAMDYGGSGVLYTTRSAV-LPESITPELOTAQ 839
Db 701 -----LQSKTLAPDKLSTSESLKKREKRDEMLGVPMRQSTIIDKEEIPBLRKLQ 755
QY 840 KVGDENFIEKTD-----NTCGPEPLVERLEKKVEEGERTIIE--880
Db 756 NVNRDIOQR-LKNDLBEQETLGGTIMPEESAKVCLDVTYINERQOMLKKOYERLQAQA 814
QY 880 -----VEEIEBEVEKYVRKDEVTLFSLKPEGEFELORDEENF-----LRELSKEE 923
Db 815 KLOGIDIDRTVOOVNOEKOEROMKHLDTIVSSNIEILNRKLIDODOIOIHLKSTTNELKSEK 874
QY 924 MDVLDDGLKME-----ATEYEKLE 941
Db 875 LOISINLORROOLEEQYIVELSTEVQSYIX 902

```

```

1      RESULT 3
2      US-08-021-601-2
3      : Sequence 2, Application US/08021601
4      : Patent No. 5591631
5      : GENERAL INFORMATION:
6      : APPLICANT: Leppla, Stephen H.
7      : APPLICANT: Kimpel, Kurt R.
8      : APPLICANT: Nichols, Peter J.
9      : APPLICANT: Arora, Naveen
10     : APPLICANT: Singh, Yogendra
11     : TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
12     : TITLE OF INVENTION: RELATED METHODS
13     : NUMBER OF SEQUENCES: 12
14     : CORRESPONDENCE ADDRESS:
15     : ADDRESSEE: Needle & Rosenberg, P.C.
16     : STREET: 133 Carnegie Way, Suite 400
17     : CITY: Atlanta
18     : STATE: Georgia
19     : COUNTRY: USA
20     : ZIP: 30303
21     : COMPUTER READABLE FORM:
22     : MEDIUM TYPE: Floppy disk
23     : COMPUTER: IBM PC compatible
24     : OPERATING SYSTEM: PC-DOS/MS-DOS
25     : SOFTWARE: Patentin Release #1.0, Version #1.25
26     : CURRENT APPLICATION DATA:
27     : APPLICATION NUMBER: US/08/021,601
28     : FILING DATE: 19930212
29     : CLASSIFICATION: 514
30     : ATTORNEY/AGENT INFORMATION:
31     : NAME: Spiatl, Gwendolyn D.
32     : REGISTRATION NUMBER: 36,016
33     : REFERENCE/DOCKET NUMBER: 1414, 057
34     : TELECOMMUNICATION INFORMATION:
35     : TELEPHONE: 404/688-0770
36     : TELEFAX: 404/688-9880
37     : INFORMATION FOR SEQ ID NO: 2:
38     : SEQUENCE CHARACTERISTICS:
39     : LENGTH: 776 amino acids
40     : TYPE: AMINO ACID
41     : TOPOLOGY: linear
42     : MOLECULE TYPE: protein
43     :
44     :
45     :
46     :
47     :
48     :
49     :
50     :
51     :
52     :
53     :
54     :
55     :
56     :
57     :
58     :
59     :
60     :
61     :
62     :
63     :
64     :
65     :
66     :
67     :
68     :
69     :
70     :
71     :
72     :
73     :
74     :
75     :
76     :
77     :
78     :
79     :
80     :
81     :
82     :
83     :
84     :
85     :
86     :
87     :
88     :
89     :
90     :
91     :
92     :
93     :
94     :
95     :
96     :
97     :
98     :
99     :
100    :

```

```
Query Match      1.8%; Score 132; DB 1; Length 776;
Best Local Similarity 17.2%; Pred.No. 0.0098;
Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31.
```

Dh	72	EMTKALIG---	KIYIDGDI	TKHISLE	ILSEDKKKIK	DIYGDALL	HEHYVAAKEGE	PV	128																			
QY	788	LSKVENSPED	IFVYKGR	NDAMDO	GGSVLY	TSRAV	LPESIT	PELOTAQAOKRG	PNT 847																			
Dh	129	L---	VIOSSE	LEYVENTE	KALN---	---VYE	IGKLS	SDILSKIN	POYKFLDYLT 176																			
QY	848	FIKDINCG---	---	PEP	LYEBLEK	VEEGERTIK	---	---	879																			
Dh	177	IKNASD	GGDILL	FTNOLK	EHPTDES	VEFLDE	ONSNEVE	FAKAPAY	IEPOHRDYLOLY 236																			
QY	879	---	---	EVEE	IEEVEK	VRDEK	EVFL	FSKLE	GEFKE----- 908																			
Dh	237	APEAFNY	MDKFNE	OEJNL	SLEELK	ODRM	SRLEKWEK	IKOHQ	KHSDSLSEEGGLKL 296																			
QY	908	---	LQDRE	NFLKEL	SKEEND	VDGLK	MEATEV	---	EKLFGALPT----- 950																			
Dh	297	QIP	EPKKD	IJHLS	QOEKEL	KRIQ	DISOS	FLTEBE	REFLKLQIDIRD	SLSEEKEL 356																		
QY	951	MAVAT	HCIF	SPCH	DRIR	FFSS	DDGIG	RLGIT	KRRNGFLKIL	EPISAD---LRTG 1006																		
Dh	357	L---	---	NR	IOVDS	SNP---	---	LSER	--EK	FLKKLIDIO	PYDINORLODTG 396																	
QY	1007	GRSRPL	---	---	---	SAFR	SGFSK	GIFDI	VP	PSKNEKEL	ELAPLLKLKG 1049																	
Dh	397	GLID	SPIN	INDV	KORYK	RDIQ	IONID	ALHOS	ISTJY	NKIY	ENNNINNLATL----- 451																	
QY	1050	VLAC	FLIV	SABAV	ALTK	CACLL	KGRI	ELAK	CIAND	PACAN	AVACIQ	TNNRDETEC 1109																
Dh	451	---	---	QAD	LVOST	---	---	---	---	---	---	459																
QY	1110	QIK	CDL	FFENS	VDE---	FNE	CAVS	RKRC	VP	PRK---SD	LEPPAD	PSVLYON	FNISDENG 1164															
Dh	459	---	---	DN	KIR	GLF	FNFK	NFKFY	SISS	NMYD	INER	ALDNE	KLRKRIQ	SLPOTR 510														
QY	1165	KWYIT	SG	INPT	FDFA	QDQL	HEH	FT	EGDN	KLV---GN	ISWR	KITL	DSG	FFTR	SAVOK	EVODP 1222												
Dh	511	AGY	LENG	---	---	---	---	---	---	---	KL	IO	RNIG	LEIK	VOV----- 534													
QY	1223	NO	GV	LVHND	NEV	LHY	ODDWY	IL	SSK	IE	NK	PEDY	IFVY	RGR	NDAM	DG	GCANVY	TR--S 1280										
Dh	534	---	---	IK	OSE	KEY	TRI	---	DA	VY	PKS	KID	IKIO	---	AO	LNI	NO	EMNKAL	GIPK	TKLIT 584								
QY	1281	SVL	PN	SIT	PEL	KAKA	SIG	DE	ST	FRT	DNT	CG	PE	PA	LY	ER	LEK	IV	EEGE	---	RIIYKEV- 1338							
Dh	585	FN	VH	NR	YAS	NIN	VE	SAV	IL	NEM	KNI	OSD---	---	LK	ITN	LY	VO	NG	EF	FTD	IFL 636							
QY	1338	---	---	BE	IEE	VE	KE	VE	KV	GR	EM	TL	FOR	LA	GE	NE	I	KO	DE	N	VRE----- 1380							
Dh	637	PN	IA	EOY	T	HODE	IE	OV	H	SG	LV	PE	RS	ITL	H	GP	SKV---	EL	RN	SE	GE	IF	HE	G	HA	V	ADY 695	
QY	1380	---	---	LS	KE	ME	---	---	FL	DE	I	KE	AS	EV	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Dh	696	AGY	LL	R	K	OS	DL	Y	T	N	S	K	K	F	I	D	I	K	E	E	B	S	N	L	725			

RESULT 4
 US-08-082-849B-2
 : Sequence 2, Application US/08082849B
 : Patent No. 5677274
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Leppert, Stephen H.
 : APPLICANT: Klimpel, Kurt R.
 : APPLICANT: Aroa, Naveen
 : APPLICANT: Singh, Yogendra
 : APPLICANT: Nichols, Peter J.
 : TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
 : TITLE OF INVENTION: Related Methods
 : NUMBER OF SEQUENCES: 35
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend and Crew LLP
 : STREET: Two Embarcadero Center, Eighth Floor
 : CITY: San Francisco
 : STATE: California

```

COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,8498
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 776 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-8498-2

```

Query Match 1.88; Score 132; DB 1; Length 776;

Best Local Similarity 17.2%; Pred. No. 0.0098;

Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

```

QY 733 ENKTVGNLSMRIRPPDGFTRSAVQKVFODPKYPGILYNHD--NEYLLYQDDWT--I 787
DB 72 EMYKAIG3--KIYIVDDITKHISLEALSEDKRKIKIDYCKDALLHEHYVAKEGEPEV 128
QY 788 LSKYENSPEYIFVYVYGRNDAMDYGGSVLYTRSAVLPESIIPELOTAAOKGRDENT 847
DB 129 L--VIOSEEDYVENTEKALN-----VYELGKILSDITSKINOPYOKELDVLTNT 176
QY 848 FIKIDNTG-----PEPPLVERLEKKVEGERTLIK----- 879
DB 177 IKNASDSGQDLFTNOIKEHPTDFSEVLEQNSNEVOEFAKAFAYVIEPHRDVLY 236
QY 879 -----EVEIEEVEKVRDKVYTLFSKLEFGKE----- 908
DB 237 APEAFNYMDKNEQFINLSLELDQRLSYEKWEKIKOHYOHWSDSLSEEGRLKLKL 296
QY 908 --LORDENFLRELSEKMDVLDGLKMEATEV---EKLFGALPT-----RKL 950
DB 297 QIPLEPKDDIHSLSQEKELKRIQIDSSDPLSTEKEFLKQIDIDISLSEKEL 356
QY 951 MAVATHTCTSCHRRIRPFSSDDOIGRLGTRKRIINGTFLIKIIPIQSAD---IKTTG 1006
DB 357 L-----NRIOVDSNP-----LSEK--EKEFLKLELDIQPYDINQRLQDDTG 396
QY 1007 GRSSRPL-----SAFRSGSGKIFDIIVPLPSKNEKELTAPLULLKLVG 1049
DB 397 GLIDSPSINDVRYKQKRDIONIDALHQSITLYIKITLYENMMNTNLTATL----- 451
QY 1050 VLACAFILVPSADAVDALKTQACLLKGRITELAKCIANPACAAVACIQTCNNRPDETEC 1109
DB 451 -----GADLVST----- 459
QY 1110 QIKGDLFENSVD--FNCVAVSRKCVPRK---SDLGEPAPDPSVLVQNFNISDENG 1164
DB 459 -----DNKIRNGIFNEKKNFKYSISNYVIYDINERPALDNERLKWRIQLSPDTR 510
QY 1165 KWTITSGINFTFAFDQLHEFHTEGDNKLY--GNISWRITKILDSGFTTSAVQKRVQDP 1222
DB 511 AGYLENG-----KILIORNIGLEIKIDVO----- 534

```

```

QY 1223 NQGVLYNHDNEYLYHQQDDWYILSSKIENKPEDYIFVYVYGRNDAMDYGGAIVYTR--S 1280
DB 534 ----IKOSEKEYIRI--DAKVPKRSKIDTKIOE-----AQNLNQMNMNALGPKTKTKLIT 584
QY 1281 SVLPNSIIELEKAAASIGRDSFTPIRNTNCGFPALVERIEKIVEGE--RIYKEV-- 1338
DB 585 FVNHNRASNIYESAVLLINEMKNQSD-----LIKVTNLYLVQNGRFVFTDITL 636
QY 1338 -----EIEIEEVEKEVKVGRTEMTLPORLAEGFNLKODENFVRE----- 1380
DB 637 PNIAGEYTHODEIYEVHSGKGLYVPSRSILLHGSKVY-ELRNDSEGFIFHFGAUVDY 695
QY 1380 ----LSKEME-----FLDEIKMEASEV 1398
DB 696 AGYLDKNOISDLYVNSKRFIDIFKEGSNTL 725

```

RESULT 5

PCT-US94-01624-2

Sequence 2, Application PC/TUS9401624

GENERAL INFORMATION:

APPLICANT: Leppla, Stephen H.

APPLICANT: Kilmpel, Kurt R.

APPLICANT: Arora, Naveen

APPLICANT: Singh, Yogendra

APPLICANT: Nichols, Peter J.

TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESS: TOWNSEND and TOWNSEND KHOURIE and CREW

STREET: Stewart Street Tower, 20th Floor, One Market

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

FILING DATE: June 25, 1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-115

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 776 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US94-01624-2

Query Match 1.88; Score 132; DB 3; Length 776;

Best Local Similarity 17.2%; Pred. No. 0.0098;

Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

```

QY 733 ENKTVGNLSMRIRPPDGFTRSAVQKVFODPKYPGILYNHD--NEYLLYQDDWT--I 787
DB 72 EMYKAIG3--KIYIVDDITKHISLEALSEDKRKIKIDYCKDALLHEHYVAKEGEPEV 128
QY 788 LSKYENSPEYIFVYVYGRNDAMDYGGSVLYTRSAVLPESIIPELOTAAOKGRDENT 847

```

```

Db 129 L---VIOSEEDYVENTEKALN-----VYEIGKILSHDLISKINCPYOKFLDLVNT 176
QY 848 FIKDNTNG-----PEPPLERLEKKVEEGERITIK-----879
Db 177 IKNASDSGDGLFTTNOLEKHPDVSVEFLQNSNEVOEPAKAFAYIEPQHRDVLQLY 236
QY 879 -----EVEIEEVEKVRDKVEVTLFSKLFEGFKE-----908
Db 237 APEAFNWKDRNEOEINSLDELQNRLSYKEMEKIKOHOYQWMSLSLEEGGLKL 296
QY 908 ---LQREENFLRELSEKEMDVLQKMEATEV---EKLFGALPT-----RKL 930
Db 297 QIPLEPKDDIHSLOCEKELKRIQDSDFLSTEKEFLKLQIDIRDSLEEEKEL 336
QY 951 MAVAFHCTSPCHDRIRFESSDDGIRGIRKIRNGFLIKLIPLOSAD---LRITG 1006
Db 357 L-----NRIOVDSNP-----LSEK--EKEFLKLLDIOPIQDINDRLQDITG 396
QY 1007 GRSSRPL-----SAFRSGFSKGFIDVPLPSKNELEKELTAPLLKLVG 1049
Db 397 GLIDSPSINDVRRQYKRDIONIDALLHQSIGSTLYNKIYEMMNINNLATL-----451
QY 1050 VLACAFILVPSADAVDAKLTACILKGCRIELACIANPACANVACLOTCNNRPDETEC 1109
Db 451 -----GADLYDST-----459
QY 1110 QIKGSDLEFNSVDE--FNECAVSRKKCVPRK---SOLGEPAPDPVLYQNMISDENG 1164
Db 459 -----DNKIRNGINTEKKNKFKISISSNIMIVDINERPALDNERLKRRIOLSPDTR 510
QY 1165 KWTYISGLNPTFDFADCOLHEFHTEGDNKLV--GNISWRIKTLDSGFTSRSAVOKFVDP 1222
Db 511 AGYLENG-----KLILQNRIGLEIKDVO-----534
QY 1223 NOPGVLYHNDIEYLYQDDWTILSKIKENKPEDYTFYRRRNDAMDGAGAVYTR--S 1260
Db 534 ---LTKOSEKEKXIRI--DAKVAPKSIDTKIOE---AOLNINQEWKALGLPYTKLIT 584
QY 1281 SYLPSIITPELEKAKSKISGRDSTFIRDNQCGEPALVERIEKTEVERGE--RIIVKEV-- 1338
Db 585 FNVHRYASNVYESAYVLLNENKKNIOSD-----LIKVTYVLDGNGRFFVFTDITL 636
QY 1338 -----EEIEEVEKEVEKVRGRTMTLFLQRLAEGFNELEKODENFVRE-----1380
Db 637 PNIAQYTHQDEIYEQVASKGLYVPSRSILHSPSKGV--ELRNDSEGFHEFHGAVDY 695
QY 1380 ---LSKEBME-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNQSDLVTKSKKFIIDIFKEEGSNL 725

```

RESULT 6
US-08-404-531B-6
Sequence 6, Application US/08404531B
Patent No. 5863724

GENERAL INFORMATION:
APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
APPLICANT: Thomas, Gilbert Cote, and Robert Gage
TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
Patent No. 5863724
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: No. 5863724 is
STREET: One Liberty Place 46th. Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

a

```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/404,531B
FILING DATE: 15-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Beardel, Lori Y.
REGISTRATION NUMBER: 34,293
REFERENCE/DOC# NUMBER: BYLR-0003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1581 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-404-531B-6

Query Match 1.7%; Score 124; DB 2; Length 1581;
Best Local Similarity 18.8%; Pred. No. 0.15; Version 1.78; Gaps 26;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

QY 868 KVEEGERITIIKEVEIEEVEKVRDKVEVTLFSKLFEGFKE--ELQREDEENFLRELSEKEMD 925
Db 480 KLSQAQRTLL-----EYSNERLQKT--NEMLRGILKLKLYAMENFCSRVKTRRK 528
QY 926 VLDGLKMEA--TEVEKLEFRALPIRKLMVAATHCTSPCHDIRIRFSSDD-----GIGRL 978
Db 529 EMTSLRAFAVYISISIFMTAIPDAVILTFV-----GHVSFFKSDSPSAVASTL 580
QY 979 GIKRRKIRNGIFLL-----KILPPIO-----SADLRITGGRSSRPLSAFRSGFSKG 1023
Db 581 SLPHILVPLPLFLSSVSVSTKALVSQKLSLELSABIREQCAPREPAPQGA-----G 636
QY 1024 IPDIVPL-----PSKNELEKELTAPLLKLVGLACAPLIVPSADAVDAKLTACILKLG 1076
Db 637 KYQAVLKVYNNRRKPARREVRDLGLQ-----RLNPSIDG--DADNFCVQIIG 684
QY 1077 C-----RIELACIANPACANVACLOTCNNRPDETECQICGLDF 1117
Db 685 FFWTDPDGIPTLSNITIRIPRGQITMIVQVCGKSSLLATLIG-----EMQKYSGAVF 738
QY 1118 ENSVVDENECYASRKKCVPRKSDIGEFPAPDP-----SYLVQNFNI--SDF 1162
Db 739 WNSLVPD-----SEGRRPQOPRAGDSGRFCQORPCGYASQXPMLNATVEENITFESPF 793
QY 1163 NGRWY-----ITSGLNPTFDFADCOLHEFHTEGDNKLVGNISWRIKT-----1205
Db 794 NKQRYKMWIEA--SLQPDIDIL-----PHGQQTQIGRGINLSIGGGRPODCREPEST 845
QY 1205 -----LDSCF-----FTSAVOKFVODPNQOPGVLYHNDIEYLYQDDWTILSS 1247
Db 846 STPMIVFLDPPFSLDVHLSHDLMQAGIILLRDKRTVAVLYTHKLOLYPHA--DWIANK 904
QY 1248 KIEKKPEDYLF--YVRGNNDAMDGAGAVYTRSSV--PNSLIPLELEKAASISGRDSTFIR 1307
Db 905 DGTIQREGITKDKQSRSCOLFHW-----KTLMNQODELER-----ETVME 946
QY 1308 TDNTGEPALVERIRKTYEGERIIVKEVEIEEVEKE 1347
Db 947 RK--APEPS--QJLPRAMSSRGULLDEDEBEBAASES 981

```

RESULT 7
US-08-404-531B-28
Sequence 28, Application US/08404531B
Patent No. 5863724
GENERAL INFORMATION:
APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
APPLICANT: Thomas, Gilbert Cote, and Robert Gage

```

TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
Patent No. 5863724
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: No. 5863724ris
STREET: One Liberty Place 46th. Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/404,531B
FILING DATE: 15-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293
REFERENCE/DOCKET NUMBER: BYLR-0003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1498 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-404-531B-28

```

```

Query Match      1.7%: Score 124; DB 2; Length 1498;
Best Local Similarity 18.8%: Pred. No. 0.14;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

```

```

QY 868 KVEBERITIEVEIEEVEKVRKDEKVTLSKLFEGEK--ELORDENETRELSEKEMD 925
DB 480 KLSQARTTL-----EYSENERLQPT--NEMLRGKIKLKLTYAMENIFCSRYEKTRRK 528
QY 926 VLDGKMEA--TEVEKLFGRALPIRKLMVAATHCFTSCHDRIFEESDD-----GIGRL 978
DB 529 EMTSIRAAVVTSTISIFNTAIPIAAVITTV-----GHVSFFKESDPSVAFASL 560
QY 979 GITRRKINGTFL-----KILPPIQ-----SADLRITGGSSRPLSAFRSGFSKG 1023
DB 581 SLFHILVPLLELSSVSVKALVSVOKLSEFLSSAEIREQCAPREPAPOGQA-----G 636
QY 1024 IEDIVPL-----PSKNEKELTAPILLKLVYLACAFILVPSADAVDAKTCACILKG 1076
DB 637 KYQAVPLVAVNRKKPARREVDLGLPLO-----RLTPSTDG-DADNFCVQIIGG 684
QY 1077 C-----RIELAKCIANPACANVACLOTCCNNRDETECOIKGDLF 1117
DB 685 FFTWPDGIPILSNITIRIPRGOLMTVGVGVCGKSSLLATLIG-----EMQKVSAAVF 738
QY 1118 ENSVYDEENECVARRKCVPRKSDLGEPADP-----SVLVQNFNT-SDF 1162
DB 739 WNSLPD-----SEGRRPDQPRAGDSGRFGCEQRPQYASQKFWLLNATVEENITFESPF 793
QY 1163 NGKVV-----INTGNTPTFADADCOLHEFHTEGDKMLVGNISMRKT-----1205
DB 794 NKQRYKWTACSLQPDIDL-----PHGDOTQLGEGKINISTGGQAPDQCPREPST 845
QY 1205 -----LDSGF-----FTRSAVQKFOVDPNQPGVLYNHDNEYLYODDWTLLS 1247
DB 846 STPIVIFLDDPFSLADVHLSDHLMQAGIELLRDKRTVVLVNHKLOYLPHA-DWIIAMK 904
QY 1248 KIENKPEYIIVYIRGRNDAMDGIIGAVVITRSSVLNPSIIPLELEKAASIGRDEFTFIR 1307

```

```

DB 905 DTIOREGTLKDFORSECOLFEHW-----KTLMNQDLEK-----ETME 946
QY 1308 TDNTCGPEPALVERLEKTEVEGERIIVKVEIEEVEKE 1347
DB 947 RK---APEPS--QGLPRAMSSRDGLLDEDEEEEAASE 981

```

```

RESULT 8
US-08-460-309-2
Sequence 2, Application US/08460309
Patent No. 5837496
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Lelivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,309
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,077
FILING DATE: 22-SEP-1993
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1130 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-309-2

```

```

Query Match      1.5%: Score 115.5; DB 2; Length 1130;
Best Local Similarity 20.0%: Pred. No. 0.46;
Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;
QY 294 NEFLHYDDWYLLSSQLEKPKDDYIFYYRGRNDAMDGYGGSVITRSPILPE-SIPN- 352
DB 6 NDVKNEDHLNGLKRIBNA-----DARNGDLRLTINDLGLKSLAIPND 49
QY 352 -----LQRAKASVGRPFNNFITDNSCGPEPPLVERLEKTADEGKLLIKRAVEIEEVEK 407
DB 50 TAAKLO-AVKDKARAND--TAKDVLAQITELHQLNDLGLKKNYKLT--ADSVAKTNAVVK 104

```



```

QY      408 EVEE-----VROEMTL-----FOLLGCFKRLDODENFEVRELSKEKEETLNECOMATE 4 458
        :   :   |   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      105 DPKNKIITADDAIVAKNLEGPADLDIKLPKIKELEDNLKNIS-ETIKELINARKOANS 153
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      459 VE-----KLFGALPRFKULMALAPHNSFLANHET-----IKRYVGS-----KLPGH 500
        :   :   |   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      164 IKAVSSSGDCIRITYKPPEIKKGYSNNIVVNKKTAIVADNLLFEGASAFIDFLAIEMRGK 223
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      501 KRFSWG-----EDYFSGIYVAKLICSSRRIPRIFRKSPRICCGIDSGLQ 545
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      224 VSFIMDVSGGVGRVEYPDLTIDDSYWRIVASRFRNGGTI-----SVR--ALDGKAS 274
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      546 LEFHGHNLSPAHSNINQWVPKGNCSCKPCKOVAlMVMWEKMGOFAKTIVAIFILSVASKA 605
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      275 IVPSTHSTSP-----PGYTILDV-----DAMALFVGGLGKL 308
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      606 DAVALDKCTCOTLKECRLBLAKCI-----SNPACAAVACLQTCNNRPDETECOICGD 659
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      309 KKADAVARIT-----FTGCGETYFDOKP-----IGLMNREREGDOCK----- 347
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      660 LFENSVDNEFCVASRKKCVPKRSD-----VGDFPV---PDBSVLVQKF-- 702
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      347 -----GCTVS-----POVEDSEGIQPDGEYALVRPIRMYPNISIVMEKERT 390
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      702 -----PMKDE-----SCGW-----FI 712
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      391 FSSSALLMYLATRIDLRDMSELDTGHIKAYSVDLGSGMASVVSNONHNDDGMKSFTLSRI 450
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      713 TRGLNPTEDAFDCOLHEFHTEENKLV---GNLSWIRITPD---GGFFT-RSAVQKF-- 762
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      451 OKONASTIVDITD-----NOENENATSSGNGNFGLDLKADKXITFGGJPLTRNLSMARP 505
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      762 -VQDERXPGLLYNHDXEYLLODYWLILSKSVENSPEDIIFYVYKGRANDAGYGGSVLV 820
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      506 EYNLKKYSGCL---KDIEISRTPYNILSS-----PDYGV-----TKGSLENVY 547
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      821 TRSANVPESIIPELOTAQOKYGRDPNTIKTDNTOG-----PEPPLYE----- 864
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      548 IVS-FPPRPFVELRVPDIVDTEINLSFTSKNESGIILLGSGTTPAPRRKRROTGOAY 605
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      864 -----RLEKKEVEEGERTTIKEVEELEEEVEKVRD-KEVTLFSKLEPGFKERORDE-E 915
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      606 VYILLNCRGLEVHLSTGARIMKLV--IRPEPNLFHDREHSVAVERIRGLFITVOVDENR 665
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
QY      914 NFLRELSEKENDVLDGLKMEATEVEKLE-GRALP 946
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Db      664 RYMQLNLTVE-----QPLEVAKLLEVGAHP 687
        :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :

RESULT          9
US-08-460-309-4
; Sequence 4, Application US/08460309
; Patent No. 5837496
GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460_309

```

```

      : CLASSIFICATION: 435
      : PRIOR APPLICATION DATA:
      :   APPLICATION NUMBER: US 08/125,077
      :   FILING DATE: 22-SEP-1993
      :   APPLICATION NUMBER: US PCT/US 94/10730
      :   FILING DATE: 21-SEP-1994
      :   PRIOR APPLICATION DATA:
      :     APPLICATION NUMBER: US 07/472,319
      :     FILING DATE: 30-JAN-1990
      :     PRIOR APPLICATION DATA:
      :       APPLICATION NUMBER: US 07/919,951
      :       FILING DATE: 27-JUL-1992
      :       ATTORNEY/AGENT INFORMATION:
      :         NAME: Campbell, Cathryn A.
      :         REGISTRATION NUMBER: 31,815
      :         REFERENCE/DOCKET NUMBER: P-LA 9721
      :         TELECOMMUNICATION INFORMATION:
      :           TELEPHONE: (519) 535-9001
      :           TELEFAX: (619) 535-8949
      :         INFORMATION FOR SEQ ID NO: 4:
      :           SEQUENCE CHARACTERISTICS:
      :             LENGTH: 3111 amino acids
      :             TYPE: amino acid
      :             TOPOLOGY: linear
      :
      : US-08-460-309-4
    
```

```

Query Match          1.5%; Score 115.5; DB 2; Length 3111;
Best Local Similarity 20.0%; Pred. No. 2.4;
Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps

QY  294 NEFLHYQDDWTILSSQIENKPPDIIFYIKRGNDAMDYGGSVITTRSPETLPE-SILPN- 352
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  1987 NDVENEDNHGLTKTRIENA-----DARNGLDLRLTLDLTGKLSAIPND 2030
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  352 ----LQRAKASVGDFNNFTITDSSCGPPEPLVERLEKTAEGEKKLLKEAVEIEEEK 407
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2031 TAAKIQA-AVKRAQAND--TAKDVLAQTTELHQNLDGLKNYNKI--ADSVAKTNAYVK 2085S
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  408 EYEK---VRDTENTL-----FORLEGFEKELQDEENFVRELSKEEKELINLOMEATE 458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2086 DPSKKIIAADDAIVNKLQEADRLIDKLPIELEDNLKNNIS-EIKLLIQARQAKNS 2144R
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  459 VE--KLFGRALPIRKLMALAPHSNFLAHET----IKYYVGS-----KLPQH 500
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2145 IKVSYSGGDCIRTKYPEIKKGSSNNIVNVWKTAVADNLFLFYLSAKETIDFAIEMRRKG 2204A
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  501 KRFSNGW-----EDYESIVYAKICSSRIPIPRFKRSPIRICGGLDSRGIQ 545
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2205 VSEFLWDVSGGVAYVEYDPLTIDSYWYRIASPTGRNGTI-----SVR--ALDGPKAS 2255S
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  546 LEFHGKHNLSPASHLNQNVPRKNGSGCKFPKDVALMWEXMGCFAPKAIAVAIFILTSVASKA 605
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2256 IYPSITHHSQSP-----PGYTTLDY-----DANAMLFVGGLTGKL 2288R
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  606 DAVIDALKITCLILEKCRLELAKCI-----SNPACANAYACLOTNNRPDETETECIOIGCD 659
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2290 KKADAVRVIT-----FTGCAGGETFYFNKP-----IGLMNFREKEGDCK----- 2328R
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  660 LPENSVVDHEHCAYSRRKKCYPRKSD-----VGDPVP--DDPSVLVQKF-- 702
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2328 -----GCTVVS-----POVEDSEGTOIEDGEGYALVSRPIRWYINISTIVEKEFT 2371R
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  702 -----DMKDF-----SGKW-----FI 712
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2372 FESSSLALLMTIALRLRDPMSEVELTDGIKYSYDLGSGMASVSYVNQHNHGKKSKSTLSLI 2433R
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  713 TRGLNPFDATPCOLHEFHTEENKLV---GNLSWRIRTPD-----GGFFT-BSAQVF-- 762
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  2432 QKQANISITYDJOT-----NQDENIATSSGNNFGDLDKADKLYFGGLPTLRNLMSWKARP 2486R
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  762 -VODPKRYGGLIYNHDNEVLLYQDDWYILTSSKVENSPEPDIFYYYKKRNDAMDYGGSVLY 820
    
```

Db 2487 EVNIAKTSGL-----KDIETSRIPYNTLS-----PDYGV-----TKGCSLENY 2528
 QY 821 TRSAVLPESTIPELOTAQAVGRDFTFIKTNTCG-----PEPLVE-----864
 Db 2529 TVS--FPKPGFVELSPVIDVGEINLSFTKNESGIIILGSGGTAPAPRRRRQGTQAV 2586
 QY 864 -----RLEKKVEEGERTIIKEVEIEEVEKVRD-KEVTLFSLFGFKELODE-E 913
 Db 2587 YVILLNGRLEVLHSTGARTMKIV--IRPEPLFHGDSVHVERTGIFTVOVDENR 2644
 QY 914 NFLRELKSEMDVLDGLKMEATEVEKLF-GRALP 946
 Db 2645 RYMONTLVE-----QPIEVKKLFVGGAP 2668

RESULT 10
 US-08-125-077-2
 ; Sequence 2, Application US/08125077
 ; Patent No. 5872231
 ; Patent No. 5872231 5840863
 ; GENERAL INFORMATION:
 ; APPLICANT: Engvall, Eva
 ; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
 ; TITLE OF INVENTION: Fragments and Uses Thereof
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell and Flores
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/125, 077
 ; FILING DATE: 22-SEP-1993
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US 94/10730
 ; FILING DATE: 21-SEP-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/472,319
 ; FILING DATE: 30-JAN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/919,951
 ; FILING DATE: 27-JUL-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Campbell, Cathryn A.
 ; REGISTRATION NUMBER: 31,815
 ; REFERENCE/DOCKET NUMBER: P-LA 9721
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9001
 ; TELEFAX: (619) 535-8949
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1130 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-125-077-2

Query Match 1.5%; Score 115.5; DB 2; Length 1130;
 Best Local Similarity 20.0%; Pred. No. 0.46;
 Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;

QY 294 NEFLHGYDDWYILSSQIEKKPDYIFVYIRGNDAWDGSGSVITRSPTLPE-SIIPN- 352

Db 6 NDVXENEDHNLGKTRILENA-----DARGDILLRTLNDITGKLSAIPND 49
 QY 352 -----LQKAKSVGRDFNNFIITDSCGPEPIYERLEKTEEEGKLLIKAVALEEVEK 407
 Db 50 TAAKIQ-AVKKRAQAND--TAKDVLAQITELHONLDGLKKYNTL--ADSVAKTAAVY 104
 QY 408 EVER--VRDTEML-----FORLEGFKELODEENFVRELSEKEKELINLOMEATE 458
 Db 105 DPSKKIILADADATVKNLEQADRLIDKPIKELEDNLKNIS-EIKELINARQANS 163
 QY 459 VE---KLEGRALPIRKLMALAPHSNLANHET-----IKYVGS-----KLPGH 500
 Db 164 IKVSVSSGGDCIRYKPEIKKGSYNNIVVWKTAVADNLLFYGSAKFIDFLAIEMRKK 223
 QY 501 KRFSMW-----EDYFSIYVAKICSSRRIPREFRSRRCGLDSRGQ 545
 Db 224 VSLMDVSGSGVGRVYPPDLIIDDSYWRIVASRTGRGT-----SVR---ALDPPKAS 274
 QY 546 LPSHGKHLSPASHINONVPRKNSGCKFPKDVALLMWEKMGFAKTAIYAFILSVASKA 605
 Db 275 IVPSTHSTSP-----PGYTIIDV-----DANAMLFVGLTGLK 308
 QY 606 DAVDALKTCTCLKECHRLAKCI-----SNPACANACIOTCNNRPDETCCQKCGD 659
 Db 309 KKADAVAYIT-----FTGCMGETYFDNKP-----IGLMWFREREKGDCK----- 347
 QY 660 LFENSVDENECASRRKCVPRKSD-----VGDEPV--PDPSTVQKF-- 702
 Db 347 -----GCTVSS-----POVEDESGTIQFDGEGYALVSPRIMWYNISTVMKRFRT 390
 QY 702 -----DMKDF-----SGKW-----FI 712
 Db 391 FSSALLMYLATRLDRDPMSEVLTGHIKVSYDLGSGMAVSYNQNHNGKMSFTLSRI 450
 QY 713 TRGLNPTFDADFDCOLHFEFTEENKLV-----GNLSWRIPTD--GGFT-BSAVQKF-- 762
 Db 451 OKONANISIVDIT-----NOENAIATSSGNNGLDLKADKTYFGGLPTLRLSKARP 505
 QY 762 -VODKTYPGILYHNDNEYLLYDDWYILSSKVENSPEDYIFVYIRGNDAWDGSGSVLY 820
 Db 506 EVNIAKTSGL-----KDIETSRIPYNTLS-----PDYGV-----TKGCSLENY 547
 QY 821 TRSAVLPESTIPELOTAQAVGRDFTFIKTNTCG-----PEPLVE-----864
 Db 548 TVS--FPKPGFVELSPVIDVGEINLSFTKNESGIIILGSGGTAPAPRRRRQGTQAV 605
 QY 864 -----RLEKKVEEGERTIIKEVEIEEVEKVRD-KEVTLFSLFGFKELODE-E 913
 Db 606 YVILLNGRLEVLHSTGARTMKIV--IRPEPLFHGDSVHVERTGIFTVOVDENR 663
 QY 914 NFLRELKSEMDVLDGLKMEATEVEKLF-GRALP 946
 Db 664 RYMONTLVE-----QPIEVKKLFVGGAP 687

RESULT 11
 US-08-125-077-4
 ; Sequence 4, Application US/08125077
 ; Patent No. 5872231
 ; Patent No. 5872231 5840863
 ; GENERAL INFORMATION:
 ; APPLICANT: Engvall, Eva
 ; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
 ; TITLE OF INVENTION: Fragments and Uses Thereof
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell and Flores
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA

```

      ZIP: 92122
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/125,077
      FILING DATE: 22-SEP-1993
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US PCT/US 94/10730
      FILING DATE: 21-SEP-1994
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/472,319
      FILING DATE: 30-JAN-1990
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/919,951
      FILING DATE: 27-JUL-1992
      ATTORNEY/AGENT INFORMATION:
      NAME: Campbell, Cathryn A.
      REGISTRATION NUMBER: 31,815
      REFERENCE/DOCKET NUMBER: P-LA 9721
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (619) 535-9001
      TELEFAX: (619) 535-8949
      INFORMATION FOR SEQ ID NO: 4:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 3111 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
      US-08-125-077-4

Query Match      1.5%, Score 115.5; DB 2; Length 3111;
Best Local Similarity 20.0%; Pred. No. 2.4;
Matches 163; Conservative 96; Mismatches 262; Indels 293; Gaps 43;

QY 294 NEFLHYQDDWYLLSQIENKPPDIIFYYYRGNDAMDYGGSVIYTRPTPE-SIIPN-352
DB 1987 NDVKNEDHLNGLKTRIEA-----DARNDDLKTLNDLTKGLSAIPND 2030
QY 352 ----LQKAKSVGRDENFIITDNSGPEPLVERLEKTAEBEGELIKEAVEIEEVEYK 407
DB 2031 TAAKQ-AVKOKARQAND-TAKDVLQDTTELHQLDGLKKNYKL-ADSVAKTNAVYK 2085
QY 408 EYK-----VRDIEMTL-----FQRLLEGFKLQDEENFVELSKEKEILNLOMATE 458
DB 2086 DESKKNITADADATVKNLEQEADRLIDKLPKLELDNLKNKIS-EIKELINOARKQANS 2144
QY 459 VE--KLGRAPIRKLMLALAHNSFLANHET-----IKYYVGS-----KLPQH 500
DB 2145 IKVSVSSGDDCIITYKPEIKKGSYNNIYVAVADNLFLYLGAKFIDFLAIEMRKQ 2204
QY 501 KRFSGW-----EDYFGSIYVAKICSSRRIPREFKSPICCGLDNRGLQ 545
DB 2205 VSFLLWDVSGVGRVYEPDLITIDSTWYRIVASRTGRNGT-----SVR---ALDGPYAS 2255
QY 546 LFSHGHNLSPAHNSINQNVPKNSGCKPEPKDVALMWKMGQFAKTAIVAFIISVASKA 605
DB 2256 IVPSTHSTSP-----PGYTILDY-----DANAMLFVGGITGKL 2289
QY 606 DAVDLAKTCTCTCLKRCRLAKCI-----SNPACAAVACIQCNNRPDETEQIKRGD 659
DB 2290 KKADAVRVT-----FTGCMGEITYFDNKP-----IGLWNEFKEGGDK- 2338
QY 660 LFEENVYDEFNECAVSRKKCVPRKSD-----VGDPEV---PDPSVLVQKF-- 702
DB 2328 -----GCTVS-----PQVEDSEGTIOFDGEGYALVSRPIRMVFNISTWVFKRT 2371
QY 702 -----DMKDF-----SGKW-----FI 712
DB 2372 FSSSALLMTATRLDRJDFMSVELDGHIKVSYDLGSGMASVVSNNHNDGKWKSFLLSRI 2431

```

```

QY 713 TRGLNPTFADFQCLHEFTEENKLV-----GNLSWRIRTPD-----GGEFT-BSAVQKF-- 762
DB 2432 QKANIISYDILT-----NOENIATSSGNNFGDLKADKXIYGGCLPTLRNLSMKARP 2486
QY 762 -VODPYPGILYNHNDYLLLYDDWYLLSKVNSPEDYIFYYKGRNDAMDYGGSVLY 820
DB 2487 EVNLKKYSGL-----KDIEISRTPYNILLS-----PDVGV-----TRGCSLENNY 2528
QY 821 TRSAVLPESTIIPLOTAACKVGRDFTFIKTNTG-----PEPPVE----- 864
DB 2529 TVS--PPKGFVFLSPVPIDVGTETLSFSTNKGIIILGSGTGPADPRRRKRGQAY 2586
QY 864 -----RLEKKVEEGERTIIEVEIEEVEKVRD-KEVTLFSLKEGFEQLORDE-E 913
DB 2587 YVLLNRGRLEHLSGTANTMKIV--IKPEPNLHFGREHSHVHERTRGIFTVOVDNR 2644
QY 914 NEFLRELSKEEMVLDGLKMEATEVEKLF-GRALP 946
DB 2645 RYMQNLTFE-----QPIEVKKLFYGGAP 2668

RESULT 12
US-08-568-459A-4
; Sequence 4, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelisen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-4

Query Match      1.5%; Score 115; DB 2; Length 1435;

```

Best Local Similarity 19.3%; Pred. No. 0.75;
Matches 171; Conservative 122; Mismatches 315; Indels 280; Gaps 46

OY	552	HULSAPHSINONVPGNSGCKFPDVALMWKEGKQFAKTAIVAFITLSVASKADAVDAL	61
Db	378	HILSKSEYE-TOKVPREN-----	AENLITI-----SEKNDAKAVSL 412
OY	612	KTCITCLKECRLEIAKLCSNPACANVACLOTNNRPDETCQKCGDFEENSVDEN-	671
Db	413	-----LNNCCAEYSKYCD-----CHHTTTLVSLVANGDNT-----IK-----EKREHIDLDDEFK	459
OY	671	-----EC-----AVSRKKCYVR-----	KSDVGD 669
Db	460	FGCDKNSVDITTKWECKNPILSTKIDVCVPPRROELCLGNDRIDYKMLMKREHLAI	519
OY	690	PVPDESVLVOKEDMKDFSGKWFITRGINPFEDAFDCOLHEHTEENKLVONLS-WRIKTP	748
Db	520	AIYESRIILKRYKKNKDDKE-----VCKIINKTP-----ADIRDILIGTDYW-----	561
OY	749	DGGFTTSRVAQKFOVDKPYBGLNHNHNEYILLYODDWITLSKAVENSPEDYIFYVYKGRN	808
Db	561	-NDLSNFKLVAKINTNSKY-----VHRNKNKDLFPEDEMWAKYIK-----	600
OY	809	DAMDGSGSVYTRSAYLPESITPELOTAQKVRDENTFK-TDNTCGPEPLVERL-	866
Db	600	DVWANI--SWFKOKTYCKEDDI-----ENIQFPRWSEMGDDYCOQRTKMIETLKY	650
OY	866	---EKKVEEG-----ERTIKEVEELIEEVEKVRDKEVTLFSKLFEGFKELODDE	913
Db	651	ECKEPCDDCDMKCKSNCSYKEMISKRKEEYKQAKOYEOFKGNMYKMYSEFKSIR-DE	708
OY	914	NFLRELSE--EMVDLDLKMEATEVEVEKLGRAIPIKKLMVAVNHCTPSCHDRIRPFSS	971
Db	709	VYIKKYSKCSNLNPFEDFKELHSDYK-----	NKTCMP-----744
OY	972	DDGIGRLIGITKRKINGTFLFKILRPIQASDLRTGTGRSSRPLSAFRSGFSKGFIDVLP	103
Db	744	--EYVDAPISITIRNNEQTSQEAYPE-ENTEL-AHRTETP-----SISEG-----P	784
OY	1032	SKNELKELIAPLLIKLYVLAACALIYP--SADAVDALKTACAL-LKG-CRIELACIAN	108
Db	785	KGNKEKEEDDLSKTI-----SVSENSRPETDAKDTISNLKLKGADVIMSPRAVIG	836
OY	1088	PACANVACLOTNN-----RPDETCQAKCGDLEFNSVADFEFNCAVSARK	113
Db	837	SSPDNINWTEQGNISGVNSKPLSDDVRPKRELEDQNSDESEHYVNH-----ISKSP	891
OY	1135	CYPRKPSDGEPPADPVSUVONENIS--DENKMWY-TGSLNTPDAPDCQJHEFHTBG	119
Db	892	SINNGDSGSSAIVSESSSNTGLSIDDDENGDTPFVOTDRIANTEDV-----IRKENADK	947
OY	1191	DNKLAVGNISWRITGLDSGEFTRSAVOKFVOPDNPGLVYVHNDENYLH-----YODDWY	124
Db	948	DEDEKGADEEHSHISES--LSSEPEKMLID-NBGSGNLSHBEYKETSMSDNVOOSGGI	100
OY	1245	LSSEIKENKPEDYIFYVYGRNDADWYGAGAVYTRSSVLEPNSIIPULEKAKSIGDFST	130
Db	1004	VNMVWEKELKDTL-----ENPS-----SSLDGKAHEELSELSPNLSDDQMS-	104
OY	1305	FIRINDCGEPALVERIEKVE-----ESEFRIIVAEVEI	1340
Db	1045	-----NTPGP--LDNTSEETTESINNEKVEREDERTLTYEYEDI	1084

RESULT 13
US-08-353-700-1
; Sequence 1, Application US/08353700
Date of Invention: 08/08/08

TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4

Query Match	1.5%;	Score 112.5;	DB 1;	Length 3248;
Best Local Similarity	17.7%;	Pred. No. 4.6;		
Matches 219;	Conservative 178;	Mismatches 408;	Indels 431;	Gaps 60;

```

QY 381 ERKETAEE--GKLLIKAEVEEEVEVEKVRD---TEMTLQRL-----ESFK 428
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 883 QRSKIOEDISAHQNVVAETLSALEKKEEQLINDKVEBDAEQLKSNHLEDSK 942
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 429 EL00DEENFVREISKKEEKLNL0MEATEVERLFGRALPIKRLM0APHS--NLANHE 487
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 943 EL0L0SET---LSERKEMSISLSNKRKEIELT0ENGLTEKINASIN0EKMN10KSE 998
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 488 TIKYV0SKLPGHKRSMGWEDYFGSIVAKICSSRIPIRYKRSPRICCGISGLOLF 547
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 999 SFAYIT-----DEB-----1008
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 548 SHGKHNLSPAHSIN0NVPK0NSGCKPKDVALMVWEKMG0FAKTAIVAIFILSVASKADA 607
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1008 -----EKISIELSDQYK0EKULI0QREETS-----AYEDLSQYKA-A 1046
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 608 VDAKT0TCLKRCRLELAKCISNPACANVACLOT0NNRPETEC0J0CGDLFENSVD 667
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1047 Q0KNKSKLECLN-----CTSLCKNRKNELE-0LK--EAFKHE0E 1084A
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 668 EFNECAVSRRKCVPRKSDVDFPDP0SVLQ0FKDKD0SGKWFITRGINPTFD0COL 727
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1085 FLTKLNAEER-----N0NMLE-----LEIYQALBREM 1114A
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 728 HEHTEENKLVGLNSRIRTPDGGFFTSAY0KFW0DPKYPGILYNH0NEYLLY0DDWYI 787
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1115 TD0N0NKSSEAGL0K0EIMTLKE--E0KNM0KEVND-----LLOEN0LM-----KV 1159
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 788 LSSNVE--NSPEDIYIVYKGRDADMDYG0SVLITRSANVLPE0IIP0EL0RA0K0VGD- 845
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1160 MKTH00NLESEPT-----RN-----SVKERESR0N0CNKRP0MDLEVEI0IDS 1205A
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 845 FNT-FIKTDNTGCPPEPLVERLEKKEVEGERIIKEVEIEEVEKVR-DKEVT---LF 898
DB 1206 YNAQVLE-----AMRNKELKLOESE-----KEKELQHELOTINGDETSMLQDMQ 1254
QY 899 SKLEGEKELORD-EENFL---RELKSEMD-----VLGDKMEATEVEKLFGRALPIR 948
DB 1255 SOEISGLKDEIDAEKYSIGPHELSTQNDNAHLOCSTQTTMKNLNELEKIC-EILOAE 1313
QY 949 KLMAVA-----THCFTS-----PCHDRIRFFSSDDGIGRIGTRKRKINGFTLLKI 993
DB 1314 KYELTEINDSRSCITATFRKMAEEVGLNEVKLLINDSGI-----LHGE-LVED 1363
QY 994 LPIIOSADLRTTG-----RSSRPLSAFRSGFSGKIDIVLPSKNKELKELTAPILLKL 1047
DB 1364 LP-----GGEGFGOPNQHPVS-----LAPLDESNGYEHLT---LSRKE 1399
QY 1048 VGV-----LACAFLLVPS-----ADAVDLKTCACILKGRICLAKCIAN--- 1088
DB 1400 VOMHFAELQEKFLSLQSEHKILHDQCOMSSKMSLQTYVDSLKAENLVLTSLNENFGD 1459
QY 1088 -----PACANAVACIOTCNRPDETECQIKGDL-----FENSVYDE 1124
DB 1460 LVKEQOLGEBGLVPSLSSCVPDSSSLSGDSSFYRALLQDTGDMSLNLNLEGAV--S 1517
QY 1125 FNECAVSRKKCVPRKSDGEPFAPDPVLYONFNISDFNG-----KMYITSGLNTPDA 1178
DB 1518 ANQCVDEVFCSSLOTYYDSLKAEN---LVLTSLNENFGDLVKEMQGLEGLVPSLSS 1574
QY 1179 F-----DCQLHEFITE---GDNKLVGNISWRIK---TIDSGFTT----- 1212
DB 1575 SCVPDSSSLSGDSSFYRALLQDTGDMSLNLNLEGAVSANQCVDEVFCSSLOEENLTR 1634
QY 1212 -----RSAVOKFVOPDNQGVLYNHNDEVLHAYODDWYIILSSKIE 1230
DB 1635 KETPAPKAVEEELSLCEVYROSLEKLEKESOGIKKNEIOLE-----QLTSS--E 1687
QY 1251 NKPEDIYFVYVYGRNDAMDYGAVVYTRSSYL-----PNSIIPLEKAKASI-GRDFS 1303
DB 1688 RQELDLCKRQYVSENEQOQKITSTLEMESKLAEKQOTDOLSTLEVALQLOGLDLS 1747
QY 1304 TFRIDNTGCPPEALVERIEK-----TVEGERIIVKEVEEI-----EEVEKEVEKV 1351
DB 1748 S--RSLGIDIEDAIOGRNESCDISKHEHTSETTERPKHDVHQIDCKDAQODLNDIKI 1805
QY 1352 GMT-----EMTLFORLAEGFNEI-----KQ 1371
DB 1806 TETGAVKFTGECGQSPDITNEPPGEDKT--QGSSECISELSPSGNALVPMDFLNGOE 1863
QY 1372 DEENE--VRELKSEMEFLDEIKMEASEVEKLFGR 1404
DB 1864 DIHNQOLVKETSNENLRLHVIEDRDKRVESLNE 1899

```

RESULT 14
PCT-US95-16216-1

```

: Sequence 1, Application PC/TUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yen, Timothy J.
: APPLICANT: Ratner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700
: FILING DATE: 09-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-16216-1

Query Match 1.5%; Score 112.5; DB 3; Length 3248;
Best Local Similarity 17.7%; Pred. No. 4.6;
Matches 219; Conservative 178; Mismatches 408; Indels 431; Gaps 60;

QY 381 EREKTAEE--GEKLLIKEAVEIEEVEKEVEKVRD---TEMTLFORLL-----EGFK 428
DB 883 QRSIKLOEDISAQNVAAETLSALENKEKELOINDKVTEBOAEIJOELKSSHLLSDSK 942
QY 429 ELQODEENFVRLSKKEKEILNLOMATEVEKLFGRALPIRKLMAALPHS-NFLANHE 487
DB 943 ELQILSBT---LSLEKREMSIISLNKREIELELQENGLIKENASLNQEMNLIQSE 998
QY 488 TITYVYGSKLPCFKRSWGMEDYFGSIYVAKICSSRRIPRFRKSPRICCGIDSGLOLF 547
DB 999 SFANYI-----DER----- 1008
QY 548 SHGKHNLEPHS:NONVPGNSGCKFPKDVALLMWEKMGQPAKTAIVATFILTSVASKADA 607
DB 1008 -----EKISLSDQYKOEKILILQRCETEN-----ATEDLSQYKKA-A 1046
QY 608 VDALKTCOTLCKEKRLELAKCISNPACANVACLOTNNRPDETECQIKGDLFENSVYD 667
DB 1047 QEKNSKLECLLSE-----CTSLCEKRNKLE-QLK--EAFKRNQE 1084
QY 668 FNECAVSRKKCVPRKSDVGDFPDPVSVLQKFDKDFSGKWFITRGINPTFDAPDQOL 727
DB 1085 FLTKLFAEER-----NONLME-----LETVOQLRSEM 1114
QY 728 HEHTEENKLVGNLSWRIRTPDGGFTFRSAVOKFVODPKYPIILYHNDEYLLYODDWYI 787
DB 1115 TDNONNSKSEAGGLKQELMTLKE--EQNNKQEVND-----LLOEEOQLM-----KV 1159
QY 788 LSSKVE--NSFEDIYFVYVYGRNDAMDYGGSYLYRSYAVLPESIIPELOTAAQKVRGD- 845
DB 1160 MKTRHEQNLSEPI-----RN-----SVKHEESRNOCNRPQMDLVEKLEISDS 1205
QY 845 FNT-FIKTDNTGCPPEPLVERLEKKEVEGERIIKEVEIEEVEKVR-DKEVT---LF 898
DB 1206 YNAQVLE-----AMRNKELKLOESE-----KEKELQHELOTIRIDLETSMLODMQ 1254
QY 899 SKLEGEKELORD-EENFL---RELKSEMD-----VLGDKMEATEVEKLFGRALPIR 948
DB 1255 SOEISGLKDEIDAEKYSIGPHELSTQNDNAHLOCSTQTTMKNLNELEKIC-EILOAE 1313
QY 949 KLMAVA-----THCFTS-----PCHDRIRFFSSDDGIGRIGTRKRKINGFTLLKI 993

```

```

Db 1314 KYELVLELNDSECTITATRKAAEEVCKLINEVKKIINDSGL-----LHGE-LVED 1363
QY 994 LPEIOSADLRTTGG-----RSSRPLSAFRSGSGKIFDIIVPLPSKNEKELTAPLLKL 1047
Db 1364 LP-----GGEFGDEQPNQHPYS-----LAPIDSENSVNEHL--LSQKE 1399
QY 1048 VGV---LACAFILIVPS-----ADAVDAIKTACCLIKGCRILACIAN---- 1088
Db 1400 VQMHFAELQEKFLSLQSEHKILHDHCQMSKSKSEQLQIVDSIKAKENVLSTNLNFOGD 1459
QY 1088 -----PACAAVACLOTQNNRPDETCQIKCGDL-----FENSVDYE 1124
Db 1460 LYKEMOLGLEEGVLPSELSSCVPSDSSLSLGDSSYRALLLEOTGDMSLSLNEGAV--S 1517
QY 1125 FNECAVSRKKCVPRKSDLEGFAPDPSVLVONENISDENG-----KWTYTGILNPTDA 1178
Db 1518 ANQCSYDEVFCSSLQTYVDSIKAKEN--LVSTNLNFOGDLYKEMOLGLEEGVLPSSLS 1574
QY 1179 F-----DCQLHEHTF--GDNKLVGKISMRK-----TIDSGFT----- 1212
Db 1575 SCVPDSSSLSSLDSSFYRALLLEQTDMSLSLNEGVSANQCSYDEVFCSSLQGEENTLR 1634
QY 1212 -----RSAYOKFVQDPNPGVLVYNDHNEYLYQDDWYIISKIE 1250
Db 1635 KETPSAPAKGVLESLICEVYRQSLKLEKMSQGIKMKKEIQELE-----QLSS--E 1687
QY 1251 NKEDYIFVYRGRNDAMDQYGGAVVYTRSSVL-----PNSITPLEKAANSI-GRDS 1303
Db 1688 RQELDLRKQYLSENQWOKLTSTVLEMSKLAARKQTEQLSTLEVARQLQGLDLS 1747
QY 1304 TFIRTDNTGCPPEALVERIEK-----TVEBERIYVEVEEI-----EEVEKEVEKY 1351
Db 1748 S--RSLIGIDTEALQGRNESCDSKEHTSETTERPKADVHQCIDKDAQDUNLDIEKI 1805
QY 1352 GRT-----EMTLFORLAEGFNEI-----KO 1371
Db 1806 TETGAVKPTGECGSEQSPDINTEPPGEDKT--QGSSECELSLSEFSGNALVPMFLNGE 1863
QY 1372 DEENF---VRELKEMEFLDEIKMASVEYKLFGR 1404
Db 1864 DIHNLQLRVKEISNENLRHLVIEDRDRVESELNE 1899

```

RESULT 15

US-08-603-753D-4

Sequence 4, Application US/08603753D

Patent No. 5891857

GENERAL INFORMATION:

APPLICANT: HOLT, JEFFREY T.

APPLICANT: JENSEN, ROY A.

APPLICANT: PAGE, DAVID L.

APPLICANT: KING, MARY-CLAIRE

APPLICANT: SZABO, CELILA I.

APPLICANT: JETTON, THOMAS L.

APPLICANT: ROBINSON-BENION, CHERYL L.

APPLICANT: THOMPSON, MARILYN E.

TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2

TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON

NUMBER OF SEQUENCES: 29

CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.

CORRESPONDENCE ADDRESS:

ADDRESSEE: ARLES A. TAYLOR, JR.

STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER

CITY: DURHAM

STATE: NORTH CAROLINA

COUNTRY: USA

ZIP: 27707

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage

OPERATING SYSTEM: Windows 3.1

SOFTWARE: WORD PERFECT 6.1 and ASCII

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/603,753D
FILING DATE: 20 FEB 1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 08/373,799
FILING DATE: 17 JAN 1995
ATTORNEY/AGENT INFORMATION:
NAME: ARLES A. TAYLOR, JR.
REGISTRATION NUMBER: 39,395
REFERENCE/DOCKET NUMBER: 1242/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 493-8000
TELEFAX: (919) 419-0383
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-08-603-753D-4
Query Match 1.5%; Score 111.5; DB 2; Length 3418;
Best Local Similarity 17.8%; Pred. No. 6.1;
Matches 275; Conservative 212; Mismatches 576; Indels 483; Gaps 72;
QY 3 LSLHTVFLCKEALNLNARSPCNERFRHSQOPT--NIIMKIRSNMNYGFSRLFTSY 59
Db 540 LEHTVVSQKEDSL-----CPNLIDNGSWPATYQNSVALK--NAGLISLKKTKK 589
QY 60 KTSFSDSSHCK-----DKSOICSIDTSFEIEQRFDLKRGWTLILERQWFOIOLATV 112
Db 590 FIVAIHDETYFKKPKIKDQKSELINCSAQF-EANFE--APLIFANADSLHLSVK 644
QY 113 LVCTFVIVPRYDAVDAIKTACCLIKGCRILACIANPSCANVACLOTQNNRPDETECO 172
Db 645 RSCS-----ONDESEPTLSLTSRFGTLIRKCSRNFTCSNNTVYSODLDYK--EAKCN 694
QY 173 IKCGDLFENSVD-----QFNECAVSRKKCVPRKSDVGEPPV-----PDRNAVYQNFNNK 222
Db 695 KKKQLQFTPEADSLSLQEGQCCENDPKS--KKVSDIKKEVLAACHPVOHSKVE-YSDT 751
QY 223 DFGSKWYI-----TSGINPTFDADFQCL-----HFFHMEKDLVGNLTWRIKTLDG 268

```

```

Db 752 DFOSOKSLLYDHENASTLLPTLSKOVLSNLVMSRGESEYMSDKLGN----- 802
QY 269 GFETRSAVOTFVODPDLPGALYNHNEFLAHYODDWYLLSSQJEN-----KPDYIFVYRG 324
Db 802 -----NYSDEVLITNIMERNQ-----DYCALNENKNVELLPREYIMVAPS 846
QY 325 RNDAMDGYSVYITRSPILPESILPNOLOKAKSVGR-----DENFI----- 368
Db 847 R-----KYOFNONTNL--RVIOKNOEFTSISKITVNDPSELESDNENNHFVOYA 895
QY 368 -----TTDSCGPEPLV--ERLEKTAEBEK-----LIKEVELEEEVE 406
Db 896 NERNNLALNTRKELHETDLTCVNEPLFKNSTWLYGDGDKOATQVSIKKDLVYLAEN 955
QY 407 KEVEKVRDTEMLFORL-----LEGFELOODEEN-----FVRELK 443
Db 956 KNSVK-QHAKMTLGDLKSDISLNDKIREKNNDYNNKAGLGPISNHSFGSFRITASN 1014
QY 444 EKEEI-----LNELOMATEVERLFGRALPIRLRLMALA--PHSNFLANHETIKYVGSKL 498
Db 1015 KEIKISEHNHKKSKMFEDIEQYPTSLACVEIIVTLALDNOKLSKPSI-----NTVS 1069
QY 499 GHRFESMGWEDYFGSIYVAKICSSRIIPRYFKSRPICCGLDGRLOLFS-----HGKHL 554
Db 1070 AHLQ-----SSVYVSDCKNSHITPO-----MLFSKODFNSNHL 1103
QY 555 SP-----AHSINONVPKNSGCKFPKDVALMWERKGOFAKTAIVAFILSVAS--KADAV 608
Db 1104 TBSOKAETTELSTILESSQF-----EFTOFKRP--YILOKSTFEVPEKOM 1149
QY 609 DALKTCTCLKECRLIACISNPACANVACLOTNNRPDETCQIKCGDLFENS--- 665
Db 1150 TILKITS--EECRADDLHVMN--APSIGVDSCKOFEGETVEIKRFPAGLLKNDCKNS 1203
QY 665 ----VDE-----FNECAVSRKKCVPRKSDVGDVFPYDPSPVLYOKFDMKD 705
Db 1204 ASGYLDENEVEGRFYSAHGTLKLVNSTALQAVKLFSDIENIS--EITSAEVHPISLSS 1262
QY 706 FSGKWFITGLNPTFPAFCOLHEFT--EENKLVNLSWRIRPDGGEFTRSAYOKFV 762
Db 1263 -----SKCHDSVSMFKIENHNDKTVSEKNNKCOLLONNIEMTGTTF-----VEEIT 1310
QY 763 QDKYPIIILYNDHEYLLODDWYLLSKYVENSPEYIFVYKGRNDAMDGYSVLYTR 822
Db 1311 EN--YKRNTEENEDNKYTAASRNSHNLFPDGSCKNDYVCIHKDETD-----LFTFD 1360
QY 823 SAVLPESIPELOTAOKYGRDNFTIKTD-----NTCGPEPLVER 864
Db 1361 Q-----HNICKLSGQPMKEG--NTQIKEDLSDLFEVAKAOEACHGNTSNKEOLTATK 1413
QY 865 LEKKVEGEERTIIEVEEIEBEYKVRDEVT--LFSKLFEGFEKLODEENFL--- 917
Db 1414 TEON-----IKDFTSPTFOTASGKNISYAKELFNKIYFPOKPELHNEFSLNSE 1465
QY 917 --BELSKEENDVLGLKMEATEV--EKLFGRALPIRKLMANVAHCHFTSPCHDRIRPFSSD 972
Db 1466 LHSIDIRKNMDI--LSYEETDIYKHKILKESYV----- 1498
QY 973 DGIGRLGIT-----RKRINGTFL-----KILPPIOSAD---LRTTGGRSSRPL 1013
Db 1498 -GTGNOLVTFQOGPERDEKIKEPTLLGFHTASGRKVKIAKESLADKVKNLFEDEKOGTSEI 1556
QY 1014 SAFRSGFSKGFIDVLPSPSNEIKELTAPLLKLVGLACAFIIVPSADAVALKTCACL 1073
Db 1557 TSFSHOWAK-----TLKYEACKDLF-----LACETIETIAARKCKEMON--SL 1598
QY 1074 LKGCRIELACIANPACANVACLOTN-----NRDET--ECOIKC--- 1114
Db 1599 NNDKNLVSIETVVPKLLSNLCROTENLTKTSKSIPLKYVHENEVEKETAKSPATCYTNO 1658
QY 1114 --GDLFENSVDDEFNECAVSRKKCVPRKSDLGEPAPDPSVLVONFNISDFNGKWIITSG 1171
Db 1659 SPYSVIENSALAFYTC--SRKTSVSQTSLL-----EAKKWLREG 1696

```

```

QY 1172 LNPTEAFDQJHEFHTEGDNKLVGNI SWRIKTLDSGFEFTRSAYOKFVODPNOGVLYNH 1231
Db 1697 I-----FDQGPB-----RINTAD-----YVGNVYLXENNSNSTIAEN 1727
QY 1232 DNEYLHQQDDWYLLSKYIENKPEYIFVYIRGRNDAMDGAGAVVTRSSVLPNSIIPBL 1291
Db 1728 DKMHLSEKODTYLSSNSMSNS-----YSHSDEVYNDSG---YLSKNNKLDGIEPVL 1776
QY 1292 EKAASISGRQPSIFIRPDNTCGPEPALVERIEKTEVEGERIIVKEV 1337
Db 1777 KNVEDQKNTSFSKYVISNVKDNANVPOTV-----NEDICVEEL 1813

```

Search completed: October 14, 1999, 00:47:20
Job time: 771 sec

